Path statistics, memory, and coarse-graining of continuous-time random walks on networks

Michael Manhart,[1,2,a)] Willow Kion-Crosby,[1] and Alexandre V. Morozov[1,b)]

[1]*Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA*
[2]*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA*

Continuous-time random walks (CTRWs) on discrete state spaces, ranging from regular lattices to complex networks, are ubiquitous across physics, chemistry, and biology. Models with coarse-grained states (for example, those employed in studies of molecular kinetics) or spatial disorder can give rise to memory and non-exponential distributions of waiting times and first-passage statistics. However, existing methods for analyzing CTRWs on complex energy landscapes do not address these effects. Here we use statistical mechanics of the nonequilibrium path ensemble to characterize first-passage CTRWs on networks with arbitrary connectivity, energy landscape, and waiting time distributions. Our approach can be applied to calculating higher moments (beyond the mean) of path length, time, and action, as well as statistics of any conservative or non-conservative force along a path. For homogeneous networks, we derive exact relations between length and time moments, quantifying the validity of approximating a continuous-time process with its discrete-time projection. For more general models, we obtain recursion relations, reminiscent of transfer matrix and exact enumeration techniques, to efficiently calculate path statistics numerically. We have implemented our algorithm in PathMAN (Path Matrix Algorithm for Networks), a Python script that users can apply to their model of choice. We demonstrate the algorithm on a few representative examples which underscore the importance of non-exponential distributions, memory, and coarse-graining in CTRWs. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4935968]

I. INTRODUCTION

We can model many dynamical systems in physics, chemistry, and biology as random walks on discrete state spaces or network structures. For example, random walks can represent proteins folding on a coarse-grained network of conformational states,[1,2] particles diffusing in disordered, fractal-like media,[3,4] populations evolving in DNA or protein sequence space,[5,6] and cells differentiating across epigenetic landscapes of regulatory states.[7,8] One can also use random walks to probe the structure of empirical complex networks, such as protein-protein interaction networks or the World Wide Web.[9–11] The central problem in these models is characterizing the statistical properties of paths taken by the system as it evolves from one state to another, especially for systems out of equilibrium. This entails understanding not only the distribution of lengths and times for these paths but also their distribution in the state space, which may reveal bottlenecks and indicate the diversity of intermediate pathways.

There is extensive literature for random walks on lattices,[4,12,13] fractals,[3,4] and random and scale-free networks[9,11,14] in the absence of an energy landscape or other objective function. Much of this work has focused especially on the scaling behavior of first-passage times and the mean square displacement, the latter being important to identifying anomalous diffusion.[3] More complex models, especially those

with energy landscapes derived from empirical models or experimental data, generally require numerical approaches. One such approach is transition path theory,[15] which relies on numerical solutions of the backward equation. This technique has been used to study Markov state models of molecular kinetics such as protein folding.[1,2]

Standard transition path theory, however, is not applicable to general continuous-time random walks[13] (CTRWs) where states may have non-exponential waiting time distributions, nor does it address the complete distribution of first-passage times beyond the mean. These problems are important in many systems. For example, molecular Markov state models require grouping large numbers of microscopic conformations of molecules into a small number of effective states;[16] the stochastic dynamics are then analyzed on this effective model.[1,2] However, this coarse-graining is known to lead to qualitative differences with the underlying microscopic dynamics.[16] In particular, the loss of information due to coarse-graining can lead to the appearance of memory, manifested as non-exponential waiting time distributions, in the coarse-grained states. Indeed, there is evidence of non-exponential distributions of time in protein conformation dynamics[17,18] and enzyme kinetics.[19,20] Non-exponential distributions can also arise from spatial disorder, as in glassy systems.[21,22] Other linear algebra-based methods besides transition path theory have been developed to treat general CTRWs,[23–25] but such methods are complicated and provide relatively little physical insight.

An alternative, more intuitive approach to CTRWs uses the path representation: statistical properties of CTRWs are

a)Electronic mail: mmanhart@fas.harvard.edu
b)Electronic mail: morozov@physics.rutgers.edu

decomposed into averages over the ensemble of all possible stochastic paths through state space. Some analytical results with this approach for arbitrary energy landscapes and waiting time distributions have been obtained, but only for 1D lattices, due to the difficulty of enumerating paths.[26–28] On the other hand, path sampling methods[29] are able to treat arbitrary network topologies, but these methods have not been developed for non-exponential waiting time distributions, and in any case, sampling is likely to be inefficient for calculating higher moments of path statistics, which are crucial when non-exponential distributions are expected.

Here we develop a generalized formalism for the path ensemble of a CTRW on a network of discrete states, regardless of their connectivity, energy landscape, or intermediate state waiting time distributions. In Sec. II we use statistical mechanics of the nonequilibrium path ensemble for a CTRW to obtain expressions for arbitrary moments of path statistics including path length, time, action, and any conservative or nonconservative force along a path. We use this formalism to deduce general relationships among the distributions of path length, time, and action, as well as several exact relationships for the case of homogeneous networks. In Sec. III we derive recursion relations, reminiscent of transfer matrix and exact enumeration techniques, to efficiently calculate various path statistics numerically, including distributions of paths in the state space. We have implemented our approach in a user-friendly Python script called PathMAN (Path Matrix Algorithm for Networks), freely available at https://github.com/michaelmanhart/pathman, that users can apply to their own models.

In Sec. IV we demonstrate the numerical algorithm on a few examples. After illustrating some basic concepts on a simple 1D random walk, we apply our method to a 1D comb to show how coarse-graining can lead to the appearance of memory, in the form of non-exponential waiting time distributions. We quantify the effect of the memory on the distribution of total path times. We further demonstrate the effect of coarse-graining in a 2D double-well potential, from which we deduce some general properties of memory arising from coarse-graining. Finally, we use our method to show how spatial disorder can also lead to non-exponential distributions of path statistics in the 2D random barrier model.

II. DISTRIBUTIONS IN THE PATH ENSEMBLE

A. Continuous-time random walks and memory

Consider a stochastic process on a finite set S of N states: the process makes discrete jumps between states with continuous-time waiting at each state in between jumps. Such a process is known as a CTRW,[13] and it can describe many physical or biological systems, such as a protein traversing a coarse-grained network of conformations toward its folded state[1,2] or a particle traveling through a disordered material.[3,4] The time the system waits in a state σ before making a jump to σ' is distributed according to $\psi(t|\sigma \rightarrow \sigma')$. In many models, this distribution depends only on the current state σ and not on the destination σ', so that $\psi(t|\sigma \rightarrow \sigma') = \psi(t|\sigma)$; such waiting time distributions are known as "separable."[30] We will

mostly assume separable distributions throughout this paper. However, since non-separable distributions arise crucially in coarse-grained models, we will also briefly discuss how to extend our results to the non-separable case. Let the raw moments of the waiting time distributions be denoted as

$$\theta^{(n)}(\sigma) = \int_0^\infty dt\, \psi(t|\sigma)\, t^n. \tag{1}$$

We assume every state σ has at least one finite moment for $n > 0$; the zeroth moment is always $\theta^{(0)}(\sigma) = 1$ by normalization. In the special case of a discrete time process, $\psi(t|\sigma) = \delta(t - \theta^{(1)}(\sigma))$ and the moments are $\theta^{(n)}(\sigma) = (\theta^{(1)}(\sigma))^n$.

Given the system has finished waiting in σ and makes a jump out, the probability of jumping to σ' is given by the matrix element $\langle \sigma'|\mathbf{Q}|\sigma \rangle$, where $\mathbf{Q}$ is an $N \times N$ matrix and $|\sigma\rangle$ denotes an N-dimensional vector with 1 at the position corresponding to the state σ and 0 everywhere else. The jump probabilities out of each state σ are therefore normalized according to $\sum_{\sigma'}\langle \sigma'|\mathbf{Q}|\sigma \rangle = 1$, with $\langle \sigma|\mathbf{Q}|\sigma \rangle = 0$ by definition (since a jump must leave the current state). The matrix $\mathbf{Q}$ imposes a network structure over the states in S, with edges directed and weighted by the entries in $\mathbf{Q}$. We can think of the jump process alone as a discrete-time projection of the model, since it describes the system's dynamics if we integrate out the continuous waiting times.

An ordinary Markov process is a special case of the above CTRW construction. A continuous-time Markov process is typically defined by a rate matrix $\mathbf{W}$ such that in a small time interval Δt, the probability of making a jump $\sigma \rightarrow \sigma'$ is $\langle \sigma'|\mathbf{W}|\sigma \rangle \Delta t$. Therefore, the probability of making the jump $\sigma \rightarrow \sigma'$, given that the system makes any jump out of σ during Δt, is

$$\frac{\langle \sigma'|\mathbf{W}|\sigma \rangle \Delta t}{\sum_{\sigma''}\langle \sigma''|\mathbf{W}|\sigma \rangle \Delta t} = \frac{\langle \sigma'|\mathbf{W}|\sigma \rangle}{\sum_{\sigma''}\langle \sigma''|\mathbf{W}|\sigma \rangle} = \langle \sigma'|\mathbf{Q}|\sigma \rangle, \tag{2}$$

which defines the relation between the Markov rate matrix $\mathbf{W}$ and the jump matrix $\mathbf{Q}$. The probability per unit Δt of waiting time $t = M\Delta t$ in σ and then making a jump out is given by

$$\frac{1}{\Delta t}\left(\sum_{\sigma'}\langle \sigma'|\mathbf{W}|\sigma \rangle \Delta t\right)\left(1 - \sum_{\sigma'}\langle \sigma'|\mathbf{W}|\sigma \rangle \Delta t\right)^M. \tag{3}$$

The waiting time distribution $\psi(t|\sigma)$ is then the continuous limit of Eq. (3),

$$\psi(t|\sigma) = \lim_{\Delta t \to 0}\left(\sum_{\sigma'}\langle \sigma'|\mathbf{W}|\sigma \rangle\right)\left(1 - \sum_{\sigma'}\langle \sigma'|\mathbf{W}|\sigma \rangle \Delta t\right)^{t/\Delta t}$$

$$= \frac{1}{\theta^{(1)}(\sigma)}e^{-t/\theta^{(1)}(\sigma)}, \tag{4}$$

where

$$\theta^{(1)}(\sigma) = \left(\sum_{\sigma'}\langle \sigma'|\mathbf{W}|\sigma \rangle\right)^{-1} \tag{5}$$

is the mean waiting time in σ. Hence, waiting times in a Markov process always have an exponential distribution. The higher moments of exponential waiting times are completely determined by the mean: $\theta^{(n)}(\sigma) = n!(\theta^{(1)}(\sigma))^n$.

In general, processes with exponential distributions of times $p(t)$ are important because they are *memoryless* in the

following sense: the probability of taking at least time t, given the process has already taken at least time t_0, is the same as taking at least t in the first place. That is, the system "forgets" the time it has already taken. Mathematically, this means that

$$\frac{P(t+t_0)}{P(t_0)} = P(t), \tag{6}$$

where $P(t) = \int_t^\infty dt'\, p(t')$ is the complementary cumulative distribution function. The only function satisfying Eq. (6) is a simple exponential $P(t) = e^{-t/\tau}$, from which it follows that $p(t) = \tau^{-1}e^{-t/\tau}$. For waiting time distributions $\psi(t|\sigma)$, non-exponential functions are therefore indicative of memory within a state: how much longer the system tends to wait in that state depends on how long it has already waited. In contrast to ordinary Markov models where $\psi(t|\sigma)$ is always exponential, models where $\psi(t|\sigma)$ may be non-exponential are sometimes known as semi-Markov processes.

B. The ensemble of first-passage paths

We approach CTRWs using the ensemble of first-passage paths[26–29,31–33] that first reach a particular final state or a set of final states from some initial conditions. We are interested in statistical properties of this ensemble such as its distributions in length, time, and space. In addition to situations where first-passage properties themselves are of interest, first-passage paths constitute fundamental building blocks of a stochastic process since the full propagator and steady state can in principle be derived from them.[4]

Let S_{final} be the set of final states, which we will treat as absorbing ($\langle\sigma'|\mathbf{Q}|\sigma\rangle = 0$ for all $\sigma \in S_{\text{final}}$ and $\sigma' \in S$) so that the first-passage condition is satisfied. Define a path φ of length ℓ to be an ordered sequence of $\ell + 1$ states: $\varphi = \{\sigma_0, \sigma_1, \ldots, \sigma_\ell\}$. Denote the probability distribution over initial states as $\pi_0(\sigma)$. Then the probability density of starting in a state σ_0 and completing the path φ at exactly time t is given by

$$\mathcal{P}[\varphi, t] = \pi_0(\sigma_0)\left(\prod_{i=0}^{\ell-1}\langle\sigma_{i+1}|\mathbf{Q}|\sigma_i\rangle\right)$$
$$\times \left[\int_0^\infty dt_0\, \psi(t_0|\sigma_0) \int_0^\infty dt_1\, \psi(t_1|\sigma_1) \cdots\right.$$
$$\left. \int_0^\infty dt_{\ell-1}\, \psi(t_{\ell-1}|\sigma_{\ell-1})\, \delta\left(t - \sum_{i=0}^{\ell-1}t_i\right)\right], \tag{7}$$

where $t_0, t_1, \ldots, t_{\ell-1}$ are the intermediate waiting times and δ is the Dirac delta function. The probability of completing the path φ irrespective of how much time it takes is then

$$\mathcal{P}[\varphi] = \int_0^\infty dt\, \mathcal{P}[\varphi, t] = \pi_0(\sigma_0)\prod_{i=0}^{\ell-1}\langle\sigma_{i+1}|\mathbf{Q}|\sigma_i\rangle. \tag{8}$$

The time-independent path probability $\mathcal{P}[\varphi]$ is convenient because we can express many path statistics of interest as averages over this distribution, analogous to averages over the Boltzmann distribution in ordinary statistical mechanics.[29,32,33] For example, let $\mathcal{F}[\varphi]$ be a functional

that measures some property of the path φ. We use angular brackets to denote the average of this quantity over the path ensemble:

$$\langle\mathcal{F}\rangle = \sum_\varphi \mathcal{P}[\varphi]\mathcal{F}[\varphi], \tag{9}$$

where the sum is over all first-passage paths φ of any length ending at states in S_{final}. Note that the partition function of the first-passage path ensemble, $\sum_\varphi \mathcal{P}[\varphi]$, always equals 1, since the process must reach one of the final states eventually. In this manner we can calculate *nonequilibrium* (first-passage) properties of the system as *equilibrium* properties of the path ensemble, which is time-independent by construction.

C. Distribution of path lengths

The simplest path property is its length $\mathcal{L}[\varphi]$, i.e., the discrete number of jumps along the path. The mean path length is then

$$\langle\mathcal{L}\rangle = \sum_\varphi \mathcal{P}[\varphi]\mathcal{L}[\varphi]. \tag{10}$$

Functionals for the higher moments of path length are simply powers of the length functional,

$$\langle\mathcal{L}^n\rangle = \sum_\varphi \mathcal{P}[\varphi](\mathcal{L}[\varphi])^n, \tag{11}$$

and the path length probability distribution is

$$\rho(\ell) = \langle\delta_{\ell,\mathcal{L}}\rangle, \tag{12}$$

where δ is the Kronecker delta. Note that the distribution of path lengths depends only on the jump matrix $\mathbf{Q}$ and not on the waiting time distributions $\psi(t|\sigma)$, and hence it characterizes the discrete-time projection of the underlying continuous-time stochastic process. In Appendix A, we show that the distribution of path lengths $\rho(\ell)$ is typically exponential asymptotically,

$$\rho(\ell) \sim e^{-\alpha\ell/\bar{\ell}}, \tag{13}$$

where α is a constant of order 1 with respect to system size and $\bar{\ell} = \langle\mathcal{L}\rangle$ is the mean path length.

D. Distribution of path times

In contrast to the discrete length of a path, there is also the continuous time of the path that accounts for the variable waiting times at the intermediate states. The distribution of total path times (first-passage time distribution) is

$$f(t) = \sum_\varphi \mathcal{P}[\varphi, t]. \tag{14}$$

Unlike the path length distribution, the path time distribution depends on *both* the jump matrix $\mathbf{Q}$ and the waiting time distributions $\psi(t|\sigma)$. We cannot evaluate $f(t)$ for arbitrary waiting time distributions $\psi(t|\sigma)$; however, we can express its

moments as simple averages over the time-independent path ensemble using path functionals (cf. Eq. (9)). That is, using Eqs. (7) and (14), we obtain

$$\int_0^\infty dt\, f(t)\, t^n = \int_0^\infty dt\, t^n \sum_\varphi \mathcal{P}[\varphi, t]$$

$$= \sum_\varphi \mathcal{P}[\varphi]\mathcal{T}^{(n)}[\varphi]$$

$$= \langle \mathcal{T}^{(n)} \rangle, \tag{15}$$

where the functional for the nth moment of path time is

$$\mathcal{T}^{(n)}[\varphi] = \int_0^\infty dt_0\, \psi(t_0|\sigma_0) \int_0^\infty dt_1\, \psi(t_1|\sigma_1) \cdots$$

$$\int_0^\infty dt_{\ell-1}\, \psi(t_{\ell-1}|\sigma_{\ell-1}) \left(\sum_{i=0}^{\ell-1} t_i \right)^n$$

$$= \sum_{j_0, j_1, \ldots, j_{\ell-1}} \binom{n}{j_0, j_1, \ldots, j_{\ell-1}}$$

$$\times \theta^{(j_0)}(\sigma_0)\theta^{(j_1)}(\sigma_1) \cdots \theta^{(j_{\ell-1})}(\sigma_{\ell-1}). \tag{16}$$

Each summation in the multinomial expansion is from 0 to n subject to the constraint $j_0 + j_1 + \cdots + j_{\ell-1} = n$. For example, the first few moments are

$$\mathcal{T}^{(1)}[\varphi] = \sum_{i=0}^{\ell-1} \theta^{(1)}(\sigma_i),$$

$$\mathcal{T}^{(2)}[\varphi] = \sum_{i=0}^{\ell-1} \theta^{(2)}(\sigma_i) + 2 \sum_{i<j} \theta^{(1)}(\sigma_i)\theta^{(1)}(\sigma_j),$$

$$\mathcal{T}^{(3)}[\varphi] = \sum_{i=0}^{\ell-1} \theta^{(3)}(\sigma_i) \tag{17}$$

$$+ 3 \sum_{i<j} \left(\theta^{(1)}(\sigma_i)\theta^{(2)}(\sigma_j) + \theta^{(2)}(\sigma_i)\theta^{(1)}(\sigma_j) \right)$$

$$+ 6 \sum_{i<j<k} \theta^{(1)}(\sigma_i)\theta^{(1)}(\sigma_j)\theta^{(1)}(\sigma_k).$$

Note that Eq. (16) implies that if any accessible intermediate state has a divergent waiting time moment of order n, then all path time moments of order n and higher must be divergent as well.

E. Path action and a general class of path functionals

For many systems, it is important to determine whether their dynamics are highly predictable or highly stochastic; that is, whether the system is likely to take one of a few high-probability paths every time or whether there is a large number of distinct paths with similar probabilities. One way to quantify this notion uses the path action, defined as

$$\mathcal{S}[\varphi] = - \sum_{i=0}^{\ell-1} \log\langle \sigma_{i+1}|\mathbf{Q}|\sigma_i \rangle, \tag{18}$$

so that path probability is $\mathcal{P}[\varphi] = \pi_0(\sigma_0)e^{-\mathcal{S}[\varphi]}$. As in classical and quantum mechanics, paths with minimum action

dominate, while paths of large action are suppressed. Note that like path lengths, action depends only on the jump probabilities and not on the waiting time distributions.

The mean path action is the Shannon entropy of the path distribution[34] (we ignore the path-independent $\log \pi_0(\sigma_0)$ contribution from the initial condition),

$$\langle \mathcal{S} \rangle = - \sum_\varphi \mathcal{P}[\varphi] \log \mathcal{P}[\varphi]. \tag{19}$$

This is consistent with the idea that the path action distribution tells us about the diversity of paths in the ensemble: low entropy (small mean action) means that a few paths with large probability dominate the process, while large entropy (large mean action) means that a diverse collection of low-probability paths contribute. The distribution of actions around this mean may be non-trivial, however. For instance, even if the mean action is large, the variance around it could either be small (the system must traverse one of the low-probability paths) or large (the system may traverse paths with a wide range of probabilities). We can characterize the action distribution by considering its higher moments. The functional for the nth moment of path action is

$$(\mathcal{S}[\varphi])^n = \left(- \sum_{i=0}^{\ell-1} \log\langle \sigma_{i+1}|\mathbf{Q}|\sigma_i \rangle \right)^n$$

$$= \sum_{j_0, j_1, \ldots, j_{\ell-1}} \binom{n}{j_0, j_1, \ldots, j_{\ell-1}}$$

$$\times \prod_{i=0}^{\ell-1} (-\log\langle \sigma_{i+1}|\mathbf{Q}|\sigma_i \rangle)^{j_i}, \tag{20}$$

so the total moments of the path action distribution are

$$\langle \mathcal{S}^n \rangle = \sum_\varphi \mathcal{P}[\varphi](\mathcal{S}[\varphi])^n. \tag{21}$$

The action functionals (Eq. (20)) share a similar multinomial form with the time functionals (Eq. (16)). This leads us to consider a more general class of path functionals with this form. Consider a path functional $\mathcal{U}$ that sums some property over jumps in a path (or edges in the network), so that for a path φ of length ℓ,

$$\mathcal{U}[\varphi] = \sum_{i=0}^{\ell-1} U(\sigma_{i+1}, \sigma_i). \tag{22}$$

In the case of action, $U(\sigma_{i+1}, \sigma_i) = -\log\langle \sigma_{i+1}|\mathbf{Q}|\sigma_i \rangle$. Equation (22) is a discretized line integral along the path φ, which suggests thinking of U as representing a force acting on the random walker as it traverses a path. The statistics of such forces over paths are especially interesting when the force is nonconservative, i.e., the line integral $\mathcal{U}[\varphi]$ depends on the whole path φ and not just on the end points. Non-transitive landscapes or non-gradient forces with this property have been considered in evolutionary theory[35] and biochemical networks.[36] However, even for conservative forces, the distribution of the line integral $\mathcal{U}[\varphi]$ may be

non-trivial over the path ensemble if there are multiple initial and final states. The moment functionals for any such quantity are again of the multinomial form,

$$(\mathcal{U}[\varphi])^n = \sum_{j_0, j_1, \ldots, j_{\ell-1}} \binom{n}{j_0, j_1, \ldots, j_{\ell-1}} \prod_{i=0}^{\ell-1} (U(\sigma_{i+1}, \sigma_i))^{j_i}. \tag{23}$$

This suggests that methods for calculating time or action moments can be applied to any path property in this general class.

F. Path statistics on a homogeneous network

We now consider these statistics of path length, time, and action in the simple case of a network with homogeneous properties. We first assume that the waiting time distributions $\psi(t|\sigma) = \psi(t)$ are identical for all states, with raw moments $\theta^{(n)}$ and cumulant moments $\theta_c^{(n)}$. We do not assume anything about the jump matrix $\mathbf{Q}$ (i.e., the network connectivity). In Appendix B, we derive an exact relation between path length and time moments for arbitrary $\psi(t)$:

$$\left\langle \mathcal{T}^{(n)} \right\rangle = \sum_{k=1}^{n} \left\langle \mathcal{L}^k \right\rangle B_{n,k} \left(\theta_c^{(1)}, \theta_c^{(2)}, \ldots, \theta_c^{(n-k+1)} \right), \tag{24}$$

where $B_{n,k}$ are the partial Bell polynomials.[37] For example, the first few moments are

$$\left\langle \mathcal{T}^{(1)} \right\rangle = \theta_c^{(1)} \left\langle \mathcal{L} \right\rangle,$$
$$\left\langle \mathcal{T}^{(2)} \right\rangle = \left(\theta_c^{(1)} \right)^2 \left\langle \mathcal{L}^2 \right\rangle + \theta_c^{(2)} \left\langle \mathcal{L} \right\rangle, \tag{25}$$
$$\left\langle \mathcal{T}^{(3)} \right\rangle = \left(\theta_c^{(1)} \right)^3 \left\langle \mathcal{L}^3 \right\rangle + 3\theta_c^{(1)}\theta_c^{(2)} \left\langle \mathcal{L}^2 \right\rangle + \theta_c^{(3)} \left\langle \mathcal{L} \right\rangle.$$

Note that the nth time moment depends on all length moments up to n. Equation (24) holds for the cumulants $\left\langle \mathcal{T}^{(n)} \right\rangle_c$ and $\left\langle \mathcal{L}^n \right\rangle_c$ as well (Appendix B). In Appendix C, we present an alternative argument for the first two moments of Eq. (25) using the central limit theorem, and in Appendix D we study the special case when $\psi(t)$ is exponential.

As these results show, we can think of path time as a convolution between path length and the intermediate waiting times: the variation in total path times arises from both variation in path lengths and variation in the waiting times. If $\psi(t)$ is a delta function (discrete-time process), then $\theta_c^{(n)} = 0$ for $n > 1$ (no variation in waiting times), and $\left\langle \mathcal{T}^{(n)} \right\rangle = (\theta^{(1)})^n \left\langle \mathcal{L}^n \right\rangle$ exactly: path lengths and times are identical up to an overall scale. This is consistent with our previous notion that the path length distribution fully describes the discrete-time projection of the process. However, even with continuous-time distributions $\psi(t)$, the approximation $\left\langle \mathcal{T}^{(n)} \right\rangle \approx (\theta^{(1)})^n \left\langle \mathcal{L}^n \right\rangle$, and therefore the approximate equivalence of the discrete- and continuous-time processes, may still hold if the waiting times are not too broadly dispersed (so the higher moments of $\psi(t)$ are not too large). We can make this observation more quantitative by expanding Eq. (24) as

$$\left\langle \mathcal{T}^{(n)} \right\rangle = \left(\theta_c^{(1)} \right)^n \left\langle \mathcal{L}^n \right\rangle$$
$$\times \left[1 + \frac{\left\langle \mathcal{L}^{n-1} \right\rangle}{\left\langle \mathcal{L}^n \right\rangle} \frac{B_{n,n-1}\left(\theta_c^{(1)}, \theta_c^{(2)} \right)}{\left(\theta_c^{(1)} \right)^n} + \cdots \right]$$
$$= \left(\theta_c^{(1)} \right)^n \left\langle \mathcal{L}^n \right\rangle \times \left[1 + \binom{n}{2} \left(\theta^{(cv)} \right)^2 \frac{\left\langle \mathcal{L}^{n-1} \right\rangle}{\left\langle \mathcal{L}^n \right\rangle} + \cdots \right], \tag{26}$$

where

$$\theta^{(cv)} = \frac{\sqrt{\theta_c^{(2)}}}{\theta^{(1)}} \tag{27}$$

is the waiting time coefficient of variation (CV), i.e., the standard deviation divided by the mean. The CV measures the relative dispersion of a distribution; it always equals 1 for exponential distributions. Equation (26) holds for the cumulants $\left\langle \mathcal{L}^n \right\rangle_c$ and $\left\langle \mathcal{T}^{(n)} \right\rangle_c$ as well.

Equation (26) implies that path length and time moments will be approximately proportional, and hence the whole distributions should be similar, if

$$\left(\theta^{(cv)} \right)^2 \frac{\left\langle \mathcal{L}^{n-1} \right\rangle}{\left\langle \mathcal{L}^n \right\rangle} \ll 1. \tag{28}$$

The quantity $\left\langle \mathcal{L}^{n-1} \right\rangle / \left\langle \mathcal{L}^n \right\rangle$ is typically of the order of the inverse mean path length $\left\langle \mathcal{L} \right\rangle = \bar{\ell}$; in particular, this is true when path lengths have an exponential distribution, which is generally the case asymptotically (Appendix A). An important exception to this rule is if lengths have a Poisson distribution, so that $\left\langle \mathcal{L}^{n-1} \right\rangle_c / \left\langle \mathcal{L}^n \right\rangle_c = 1$ in the cumulant version of Eq. (26). Apart from this special case, the condition of Eq. (28) is equivalent to

$$\left(\theta^{(cv)} \right)^2 \ll \bar{\ell}, \tag{29}$$

that is, the waiting time distribution must be sufficiently narrow compared to the mean path length. In many cases we expect this to hold, since $\theta^{(cv)} \sim 1$ for exponential-like waiting time distributions and the mean path length $\bar{\ell}$ is usually very large. We will investigate the validity of this condition in later examples.

We also consider path action on a homogeneous network. Path action depends only on the jump matrix $\mathbf{Q}$ and not on the waiting time distributions $\psi(t|\sigma)$, so as a simple example, we take all states in S to have the same number γ of outgoing jumps (nearest neighbors on the network) and all such jumps to have equal probability γ^{-1}. Therefore, the probability of a path is $\mathcal{P}[\varphi] = \gamma^{-\mathcal{L}[\varphi]}$ and the action is $\mathcal{S}[\varphi] = \mathcal{L}[\varphi] \log \gamma$. This means that the distribution of path actions is exactly equivalent to that of path lengths (rescaled by a factor of $\log \gamma$), and the moments are

$$\left\langle \mathcal{S}^n \right\rangle = \left\langle \mathcal{L}^n \right\rangle \log^n \gamma. \tag{30}$$

Since path lengths typically have an exponential distribution asymptotically (Eq. (13), Appendix A), path action will therefore also be asymptotically exponential as well, with mean $\bar{\ell} \log \gamma$.

III. MATRIX FORMULATION AND NUMERICAL ALGORITHM

Besides gaining general insights into the relationships among distributions of path lengths, times, and actions, the path ensemble formalism is convenient because we can efficiently calculate many ensemble averages using recursion relations that implicitly sum over all paths. We now derive these relations and show how to implement them numerically.

A. Recursion relations

We reformulate the problem in terms of matrices to express the sums over paths more explicitly. Let $\mathbf{T}_\ell^{(n)}$ be an $N \times N$ matrix (N is the number of states in S) such that the matrix element $\langle \sigma' | \mathbf{T}_\ell^{(n)} | \sigma \rangle$ is the nth time moment of all paths of exactly length ℓ from σ to σ'. In particular, the zeroth-order matrix $\mathbf{T}_\ell^{(0)}$ gives the total probability of all paths going from σ to σ' in exactly ℓ jumps. The initial condition is

$$\mathbf{T}_0^{(n)} = \delta_{n,0}\mathbf{1}, \tag{31}$$

where $\mathbf{1}$ is an $N \times N$ identity matrix. If our path ensemble is the set of first-passage paths to final states S_{final} with an initial distribution vector $|\pi_0\rangle = \sum_\sigma \pi_0(\sigma)|\sigma\rangle$, then

$$\sum_{\sigma \in S_{\text{final}}} \langle \sigma | \mathbf{T}_\ell^{(n)} | \pi_0 \rangle = \sum_\varphi \mathcal{P}[\varphi] \mathcal{T}^{(n)}[\varphi] \delta_{\ell, \mathcal{L}[\varphi]} \tag{32}$$

is the nth time moment for all paths of exactly length ℓ, and

$$\sum_{\ell=0}^\infty \sum_{\sigma \in S_{\text{final}}} \langle \sigma | \mathbf{T}_\ell^{(n)} | \pi_0 \rangle = \left\langle \mathcal{T}^{(n)} \right\rangle \tag{33}$$

is the moment averaged over paths of all lengths. This expression illustrates how to express the previous path ensemble averages in the matrix formulation.

The key advantage of the $\mathbf{T}_\ell^{(n)}$ matrices is that they obey the following recursion relation (Appendix E):

$$\mathbf{T}_\ell^{(n)} = \mathbf{Q} \sum_{j=0}^n \binom{n}{j} \mathbf{\Theta}^{(j)} \mathbf{T}_{\ell-1}^{(n-j)}, \tag{34}$$

where $\mathbf{\Theta}^{(n)}$ is an $N \times N$ matrix with waiting time moments for each state along the diagonal:

$$\langle \sigma' | \mathbf{\Theta}^{(n)} | \sigma \rangle = \delta_{\sigma',\sigma} \theta^{(n)}(\sigma). \tag{35}$$

Appendix E also shows how this recursion relation for the path time moments generalizes to the case of non-separable waiting time distributions $\psi(t|\sigma \to \sigma')$. For the total probability ($n = 0$), the recursion relation of Eq. (34) is simply multiplication by the jump matrix: $\mathbf{T}_\ell^{(0)} = \mathbf{Q}\mathbf{T}_{\ell-1}^{(0)}$, since the total probability of going from one state to another in exactly ℓ jumps must be given by the product of the jump matrices $\mathbf{T}_\ell^{(0)} = \mathbf{Q}^\ell$.

Owing to the similar multinomial form of their path functionals (compare Eqs. (16) and (20)), the path action

moments obey a similar recursion relation. Define $\langle \sigma' | \mathbf{S}_\ell^{(n)} | \sigma \rangle$ to be the nth action moment of all paths of length ℓ from σ to σ', so that

$$\sum_{\ell=0}^\infty \sum_{\sigma \in S_{\text{final}}} \langle \sigma | \mathbf{S}_\ell^{(n)} | \pi_0 \rangle = \langle \mathcal{S}^n \rangle. \tag{36}$$

In Appendix E, we show that these matrices obey the recursion relation

$$\mathbf{S}_\ell^{(n)} = \sum_{j=0}^n \binom{n}{j} \tilde{\mathbf{Q}}^{(j)} \mathbf{S}_{\ell-1}^{(n-j)}, \tag{37}$$

where the matrix $\tilde{\mathbf{Q}}^{(j)}$ is defined so that

$$\langle \sigma' | \tilde{\mathbf{Q}}^{(j)} | \sigma \rangle = \langle \sigma' | \mathbf{Q} | \sigma \rangle (-\log \langle \sigma' | \mathbf{Q} | \sigma \rangle)^j. \tag{38}$$

In fact, if $\mathbf{U}_\ell^{(n)}$ is the matrix such that

$$\sum_{\ell=0}^\infty \sum_{\sigma \in S_{\text{final}}} \langle \sigma | \mathbf{U}_\ell^{(n)} | \pi_0 \rangle = \langle \mathcal{U}^n \rangle \tag{39}$$

for any path functional $\mathcal{U}$ in Eq. (22), it obeys the recursion relation

$$\mathbf{U}_\ell^{(n)} = \sum_{j=0}^n \binom{n}{j} \mathbf{\Omega}^{(j)} \mathbf{U}_{\ell-1}^{(n-j)}, \tag{40}$$

where $\langle \sigma' | \mathbf{\Omega}^{(j)} | \sigma \rangle = \langle \sigma' | \mathbf{Q} | \sigma \rangle (U(\sigma',\sigma))^j$ (Appendix E). Therefore, recursion relations of this form extend to a wide class of path statistics.

B. Transfer matrices

To calculate the nth moment of time or action, we must carry out the recursion relation of Eq. (34) or (37) for all moments up to n. We can unify all these steps into a single transfer matrix operation convenient for numerical use. Let n_{max} be the maximum moment of interest. Define the $N(n_{\text{max}} + 1)$-dimensional column vector $|\tau(\ell)\rangle$ as a concatenation of $\mathbf{T}_\ell^{(n)}|\pi_0\rangle$ for all $n \in \{0, 1, \ldots, n_{\text{max}}\}$:

$$|\tau(\ell)\rangle = \begin{bmatrix} \mathbf{T}_\ell^{(0)}|\pi_0\rangle \\ \mathbf{T}_\ell^{(1)}|\pi_0\rangle \\ \vdots \\ \mathbf{T}_\ell^{(n_{\text{max}})}|\pi_0\rangle \end{bmatrix}. \tag{41}$$

Define the basis vectors $|\sigma, n\rangle$ for $\sigma \in S$ and $n \in \{0, 1, \ldots, n_{\text{max}}\}$ so that the (σ, n) entry of $|\tau(\ell)\rangle$ is the nth time moment at state σ at the ℓth jump: $\langle \sigma, n | \tau(\ell) \rangle = \langle \sigma | \mathbf{T}_\ell^{(n)} | \pi_0 \rangle$. We similarly define the action vector

$$|\eta(\ell)\rangle = \begin{bmatrix} \mathbf{S}_\ell^{(0)}|\pi_0\rangle \\ \mathbf{S}_\ell^{(1)}|\pi_0\rangle \\ \vdots \\ \mathbf{S}_\ell^{(n_{\text{max}})}|\pi_0\rangle \end{bmatrix}. \tag{42}$$

Now define the $N(n_{\text{max}} + 1) \times N(n_{\text{max}} + 1)$ matrices:

$$\mathbf{K} = \begin{bmatrix} \binom{0}{0}\mathbf{Q}\mathbf{\Theta}^{(0)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \binom{1}{1}\mathbf{Q}\mathbf{\Theta}^{(1)} & \binom{1}{0}\mathbf{Q}\mathbf{\Theta}^{(0)} & \mathbf{0} & \cdots & \mathbf{0} \\ \binom{2}{2}\mathbf{Q}\mathbf{\Theta}^{(2)} & \binom{2}{1}\mathbf{Q}\mathbf{\Theta}^{(1)} & \binom{2}{0}\mathbf{Q}\mathbf{\Theta}^{(0)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \binom{n_{\max}}{n_{\max}}\mathbf{Q}\mathbf{\Theta}^{(n_{\max})} & \binom{n_{\max}}{n_{\max}-1}\mathbf{Q}\mathbf{\Theta}^{(n_{\max}-1)} & \binom{n_{\max}}{n_{\max}-2}\mathbf{Q}\mathbf{\Theta}^{(n_{\max}-2)} & \cdots & \binom{n_{\max}}{0}\mathbf{Q}\mathbf{\Theta}^{(0)} \end{bmatrix}, \tag{43}$$

$$\mathbf{G} = \begin{bmatrix} \binom{0}{0}\tilde{\mathbf{Q}}^{(0)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \binom{1}{1}\tilde{\mathbf{Q}}^{(1)} & \binom{1}{0}\tilde{\mathbf{Q}}^{(0)} & \mathbf{0} & \cdots & \mathbf{0} \\ \binom{2}{2}\tilde{\mathbf{Q}}^{(2)} & \binom{2}{1}\tilde{\mathbf{Q}}^{(1)} & \binom{2}{0}\tilde{\mathbf{Q}}^{(0)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \binom{n_{\max}}{n_{\max}}\tilde{\mathbf{Q}}^{(n_{\max})} & \binom{n_{\max}}{n_{\max}-1}\tilde{\mathbf{Q}}^{(n_{\max}-1)} & \binom{n_{\max}}{n_{\max}-2}\tilde{\mathbf{Q}}^{(n_{\max}-2)} & \cdots & \binom{n_{\max}}{0}\tilde{\mathbf{Q}}^{(0)} \end{bmatrix},$$

where each $\mathbf{0}$ is an $N \times N$ zero matrix. We can express the recursion relations of Eqs. (34) and (37) for all $n \in \{0, 1, \ldots, n_{\max}\}$ as

$$|\tau(\ell)\rangle = \mathbf{K}|\tau(\ell-1)\rangle, \quad |\eta(\ell)\rangle = \mathbf{G}|\eta(\ell-1)\rangle. \tag{44}$$

These recursion relations have the solutions

$$|\tau(\ell)\rangle = \mathbf{K}^{\ell}|\tau(0)\rangle, \quad |\eta(\ell)\rangle = \mathbf{G}^{\ell}|\eta(0)\rangle, \tag{45}$$

where the initial conditions are

$$|\tau(0)\rangle = |\eta(0)\rangle = \begin{bmatrix} |\pi_0\rangle \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}. \tag{46}$$

Here each $\mathbf{0}$ represents a zero column vector of length N. We can think of $\mathbf{K}$ and $\mathbf{G}$ as transfer matrices that iteratively generate sums over the path ensemble to calculate moments. This is analogous to transfer matrices in spin systems that generate the sums over spin configurations to calculate the partition function.[38] The zeroth-order version of this formalism, which simply calculates powers of the jump matrix $\mathbf{Q}$, is equivalent to the exact enumeration method for discrete-time random walks.[3,39]

We can obtain most path statistics of interest by various matrix and inner products on these vectors. Define the cumulative moment vectors:

$$|\tau\rangle = \sum_{\ell=0}^{\infty} |\tau(\ell)\rangle, \quad |\eta\rangle = \sum_{\ell=0}^{\infty} |\eta(\ell)\rangle. \tag{47}$$

Elements of these vectors are (using Eqs. (41) and (42))

$$\langle\sigma, n|\tau\rangle = \sum_{\ell=0}^{\infty} \langle\sigma|\mathbf{T}_{\ell}^{(n)}|\pi_0\rangle,$$

$$\langle\sigma, n|\eta\rangle = \sum_{\ell=0}^{\infty} \langle\sigma|\mathbf{S}_{\ell}^{(n)}|\pi_0\rangle. \tag{48}$$

These represent the total nth moments of time and action for all paths through each state σ, but weighted by the number of visits to that state since the sum over ℓ counts a path's contribution each time it visits σ. In the case of $n = 0$,

$$\langle\sigma, 0|\tau\rangle = \langle\sigma, 0|\eta\rangle = \sum_{\ell=0}^{\infty} \langle\sigma|\mathbf{Q}^{\ell}|\pi_0\rangle, \tag{49}$$

since $\mathbf{T}_{\ell}^{(0)} = \mathbf{S}_{\ell}^{(0)} = \mathbf{Q}^{\ell}$ (Eqs. (34) and (37)). This is actually the average number of visits $v(\sigma)$ to a state σ, since the probability of a path is counted each time it visits σ. For an intermediate state σ, multiplying the mean number of visits $v(\sigma)$ by the mean waiting time $\theta^{(1)}(\sigma)$ gives the average time spent in σ. When σ is a final state in S_{final}, the random walk can only visit it once (if it absorbs at that final state) or zero times (if it absorbs at a different final state), and thus the average number of visits $v(\sigma)$ equals the probability of reaching that final state σ (commitment probability).

If there are multiple final states in S_{final}, we often wish to sum path statistics over all of them. Define the N-dimensional row vector $\langle\text{final}| = \sum_{\sigma \in S_{\text{final}}}\langle\sigma|$ (with 1 at the position for each final state and 0 everywhere else) and the $(n_{\max}+1) \times N(n_{\max}+1)$ matrix

$$\mathbf{F} = \begin{bmatrix} \langle\text{final}| & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \langle\text{final}| & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \langle\text{final}| \end{bmatrix}, \tag{50}$$

where each $\mathbf{0}$ is a zero row vector of length N. Multiplying this matrix on a corresponding vector will sum over all final states for each moment, leaving an $(n_{\max}+1)$-dimensional

vector with the total moments. For example,

$$
\mathbf{F}|\tau(\ell)\rangle = \begin{bmatrix} \langle \text{final}|\mathbf{T}_\ell^{(0)}|\pi_0\rangle \\ \langle \text{final}|\mathbf{T}_\ell^{(1)}|\pi_0\rangle \\ \vdots \\ \langle \text{final}|\mathbf{T}_\ell^{(n_{\max})}|\pi_0\rangle \end{bmatrix}
$$

$$
= \begin{bmatrix} \bar{t}^{(0)}(\ell) \\ \bar{t}^{(1)}(\ell) \\ \vdots \\ \bar{t}^{(n_{\max})}(\ell) \end{bmatrix}
$$

$$
= |\bar{t}(\ell)\rangle, \tag{51}
$$

where we use the shorthand $\bar{t}^{(n)}(\ell)$ for the total nth time moment absorbed at the ℓth jump. Note that $\bar{t}^{(0)}(\ell) = \rho(\ell)$ is the probability of reaching any of the final states in exactly ℓ jumps. Thus this method automatically calculates the entire path length distribution. On the cumulative time vector $|\tau\rangle$, $\mathbf{F}$ returns the total time moments,

$$
\mathbf{F}|\tau\rangle = \begin{bmatrix} \sum_{\ell=0}^{\infty}\langle \text{final}|\mathbf{T}_\ell^{(0)}|\pi_0\rangle \\ \sum_{\ell=0}^{\infty}\langle \text{final}|\mathbf{T}_\ell^{(1)}|\pi_0\rangle \\ \vdots \\ \sum_{\ell=0}^{\infty}\langle \text{final}|\mathbf{T}_\ell^{(n_{\max})}|\pi_0\rangle \end{bmatrix} = \begin{bmatrix} \bar{t}^{(0)} \\ \bar{t}^{(1)} \\ \vdots \\ \bar{t}^{(n_{\max})} \end{bmatrix}, \tag{52}
$$

where $\bar{t}^{(n)} = \sum_{\ell=0}^{\infty} \bar{t}^{(n)}(\ell) = \langle \mathcal{T}^{(n)}\rangle$ is the total time moment over all paths. The matrix $\mathbf{F}$ similarly acts on the action vectors $|\eta(\ell)\rangle$ and $|\eta\rangle$,

$$
\mathbf{F}|\eta(\ell)\rangle = \begin{bmatrix} \bar{s}^{(0)}(\ell) \\ \bar{s}^{(1)}(\ell) \\ \vdots \\ \bar{s}^{(n_{\max})}(\ell) \end{bmatrix} = |\bar{s}(\ell)\rangle,
$$

$$
\mathbf{F}|\eta\rangle = \begin{bmatrix} \bar{s}^{(0)} \\ \bar{s}^{(1)} \\ \vdots \\ \bar{s}^{(n_{\max})} \end{bmatrix}, \tag{53}
$$

where $\bar{s}^{(n)}(\ell)$ is the nth action moment absorbed in all final states at the ℓth jump and $\bar{s}^{(n)}$ is the total nth action moment.

Finally, for any function of state $B(\sigma)$, we can calculate the average value of that function at the ℓth intermediate jump. Define two $N(n_{\max}+1)$-dimensional row vectors, one for intermediate states and one for final states:

$$
\langle B_{\text{int}}| = \left[\left(\sum_{\sigma \notin S_{\text{final}}} B(\sigma)\langle\sigma|\right) \quad \mathbf{0} \quad \cdots \quad \mathbf{0}\right],
$$

$$
\langle B_{\text{final}}| = \left[\left(\sum_{\sigma \in S_{\text{final}}} B(\sigma)\langle\sigma|\right) \quad \mathbf{0} \quad \cdots \quad \mathbf{0}\right], \tag{54}
$$

where each vector has $n_{\max}$ zero row vectors $\mathbf{0}$ of length N. Acting with the row vector $\langle B_{\text{int}}|$ on $|\tau(\ell)\rangle$ and $\langle B_{\text{final}}|$ on $\sum_{\ell'=0}^{\ell}|\tau(\ell')\rangle$ returns the value of $B(\sigma)$ averaged over the probability distribution across all states at the ℓth jump:

$$
\bar{B}(\ell) = \langle B_{\text{int}}|\tau(\ell)\rangle + \sum_{\ell'=0}^{\ell}\langle B_{\text{final}}|\tau(\ell')\rangle
$$

$$
= \sum_{\sigma \notin S_{\text{final}}} B(\sigma)\langle\sigma|\mathbf{T}_\ell^{(0)}|\pi_0\rangle
$$

$$
+ \sum_{\ell'=0}^{\ell}\sum_{\sigma \in S_{\text{final}}} B(\sigma)\langle\sigma|\mathbf{T}_{\ell'}^{(0)}|\pi_0\rangle. \tag{55}
$$

For the final states we must sum over path lengths up to ℓ to account for the total probability absorbed at final states thus far. For example, if $B(\sigma) = \theta^{(1)}(\sigma)$, $\bar{B}(\ell)$ tells us the unconditional mean time spent at the ℓth intermediate jump. If $B(\sigma)$ is set to a position in space corresponding to state σ (for systems that allow embedding of states into physical space), $\bar{B}(\ell)$ is the average position at the ℓth intermediate jump, which over all ℓ traces the average path of the system.

C. Convergence and asymptotic behavior of path sums

To numerically calculate the foregoing matrix quantities, we must truncate the sums over path lengths ℓ at some suitable cutoff Λ. If there are no loops in the network, then the jump matrix $\mathbf{Q}$ is nilpotent, meaning there is a maximum possible path length Λ such that $\mathbf{Q}^\ell = 0$ for all $\ell > \Lambda$. In this case, all sums converge exactly after Λ jumps. If the network has loops, paths of arbitrarily long length have nonzero probability. We must then choose a desired precision $\epsilon \ll 1$ and truncate the sums at $\ell = \Lambda$ when

$$
1 - \sum_{\ell=0}^{\Lambda}\rho(\ell) < \epsilon \quad \text{and} \quad \frac{\bar{t}^{(n_{\max})}(\Lambda)}{\sum_{\ell=0}^{\Lambda}\bar{t}^{(n_{\max})}(\ell)} < \epsilon. \tag{56}
$$

The first condition guarantees that the total probability has converged: all remaining paths have total probability less than ϵ. The second condition indicates that the Λth contribution to the maximum moment $n_{\max}$ is sufficiently small relative to the total moment calculated so far.

A potential problem with the second convergence condition arises when the state space is periodic, so the final states can only be reached in a number of jumps ℓ that is an integer multiple of the periodicity (plus a constant). For instance, square lattices have a periodicity of 2. In that case, $\bar{t}^{(n_{\max})}(\ell)$ will alternate between zero and nonzero values as ℓ alternates between even and odd values. To prevent these zero values of $\bar{t}^{(n_{\max})}(\ell)$ from trivially satisfying the second condition in Eq. (56), we also require that $\bar{t}^{(n_{\max})}(\Lambda)$ be nonzero. A more subtle problem can arise if there are very low probability paths with very large contributions to the higher time moments. For example, one can construct a model where there are extremely long paths with probabilities much smaller than ϵ but which make arbitrarily large contributions to the total time moments due to the waiting time moments at those states. The algorithm will satisfy the convergence criteria before these paths are summed and therefore miss their contributions. This is an extreme example, but in general one may need to reconsider the convergence test depending on the properties of the model at hand.

How does the cutoff Λ depend on the maximum moment $n_{\max}$? To address this, we must determine the asymptotic

behavior of $\bar{t}^{(n)}(\ell)$ for different n. As long as the network has loops, the path length probability distribution $\rho(\ell)$ is asymptotically exponential (Eq. (13); see Appendix A). To estimate the asymptotic dependence of the higher time moments on path length, we consider the special case of identical waiting time distributions as in Sec. II F. Since

$$\bar{t}^{(n)}(\ell) = \sum_{\varphi} \mathcal{P}[\varphi]\mathcal{T}^{(n)}[\varphi]\delta_{\ell,\mathcal{L}[\varphi]}, \qquad (57)$$

we can use the approximation $\mathcal{T}^{(n)}[\varphi] \approx (\theta^{(1)})^n (\mathcal{L}[\varphi])^n$ from Eq. (26) (valid when the waiting time distributions are not too dispersed) to obtain

$$\bar{t}^{(n)}(\ell) \approx \left(\theta^{(1)}\right)^n \sum_{\varphi} \mathcal{P}[\varphi](\mathcal{L}[\varphi])^n \delta_{\ell,\mathcal{L}[\varphi]}$$
$$= \left(\theta^{(1)}\ell\right)^n \rho(\ell)$$
$$\sim \left(\theta^{(1)}\ell\right)^n e^{-\alpha\ell/\bar{\ell}}. \qquad (58)$$

Since $\ell^n e^{-\alpha\ell/\bar{\ell}} = e^{-\alpha\ell/\bar{\ell}+n\log\ell}$, the higher moments decay nearly exponentially (up to a logarithmic correction) with the same rate as the probability, set by the mean path length $\bar{\ell}$. We expect this asymptotic behavior to remain valid even when the waiting time distributions are not all the same, as long as the length and time moments are approximately proportional; we will empirically verify this expectation in later examples.

Although all $\bar{t}^{(n)}(\ell)$ asymptotically decay with exponential dependence on ℓ, the logarithmic correction in the exponent shifts the exponential regime toward larger ℓ for higher moments; this is why we must test convergence on the maximum moment $n_{\max}$ in Eq. (56). Indeed, $\bar{t}^{(n)}(\ell)$ in Eq. (58) is maximized at $\ell_{\max} = n\bar{\ell}/\alpha$, after which exponential decay sets in. Thus, we expect scaling for the cutoff to be $\Lambda \sim n_{\max}\bar{\ell}$ to leading order.

The approximate exponential dependence of the moments also enables a convergence scheme more sophisticated than Eq. (56). We can simply calculate the moments for path lengths until all $\bar{t}^{(n)}(\ell)$ have reached their exponential tails, fit exponential functions, and then extrapolate to infer the contributions of the longer paths. Conceptually, this means that all long path behavior is contained in the statistics of shorter paths, since long paths are simply short paths with many loops.[31,32] In practice, this procedure can help to avoid calculating extremely long paths unnecessarily.

D. Numerical implementation in PathMAN

We have implemented the aforementioned matrix formulation in a Python script called PathMAN (Path Matrix Algorithm for Networks), available at https://github.com/michaelmanhart/pathman with additional scripts for generating examples and analyzing output. Figure 1 shows the pseudocode. Since the jump matrix $\mathbf{Q}$ is typically very sparse, we can store all matrices in sparse formats for efficient storage and computation using SciPy's sparse linear algebra module.[40] The script is general enough to treat any CTRW on a finite discrete space given a list of states, their jump probabilities, and at least their first waiting time moments. The current implementation assumes separable waiting time distributions, but modifying it to run the calculations for

---

Define initial distribution $|\pi_0\rangle$ and set of final states S_{final}
Define jump and waiting time matrices $\mathbf{Q}$, $\boldsymbol{\Theta}^{(n)}$ (Eq. 35), $\tilde{\mathbf{Q}}$ (Eq. 38)
Define transfer matrices $\mathbf{K}$ and $\mathbf{G}$ (Eq. 43)
Define final state sum matrix $\mathbf{F}$ (Eq. 50)
Define row vector $\langle B|$ (Eq. 54) for each state function $B(\sigma)$
Define initial transfer vectors $|\tau(0)\rangle = |\eta(0)\rangle = (|\pi_0\rangle, \mathbf{0}, \mathbf{0}, \dots, \mathbf{0})^{\mathrm{T}}$ (Eq. 46)
Define initial cumulative vectors $|\tau\rangle = |\tau(0)\rangle$, $|\eta\rangle = |\eta(0)\rangle$

For $\ell \in \{1, 2, 3, \dots\}$:
    Update $|\tau(\ell)\rangle = \mathbf{K}|\tau(\ell-1)\rangle$
    Update $|\eta(\ell)\rangle = \mathbf{G}|\eta(\ell-1)\rangle$
    Increment $|\tau\rangle$ by $|\tau(\ell)\rangle$
    Increment $|\eta\rangle$ by $|\eta(\ell)\rangle$
    Set $|\bar{t}(\ell)\rangle = \mathbf{F}|\tau(\ell)\rangle$
    Set $\bar{B}(\ell) = \langle B|\tau(\ell)\rangle$

    If $1 - \sum_{\ell'=0}^{\ell} \rho(\ell') < \epsilon$:
        If $\bar{t}^{(n_{\max})}(\ell) > 0$ **and** $\bar{t}^{(n_{\max})}(\ell)/\sum_{\ell'=0}^{\ell} \bar{t}^{(n_{\max})}(\ell') < \epsilon$:
        Break

Output distributions $|\tau\rangle$ and $|\eta\rangle$ of cumulative moments over states
Output distributions $|\bar{t}(\ell)\rangle$ and $\bar{B}(\ell)$ over path lengths

FIG. 1. Pseudocode for the matrix calculations implemented in PathMAN.

non-separable distributions (Appendix E) is straightforward. The user can specify any path boundary conditions (initial distribution and final states) and functions of state $B(\sigma)$ to average over. The script reads all input data from plain text files in a simple format (see GitHub repository for documentation).

The rate-limiting step of the algorithm is multiplying the transfer matrices $\mathbf{K}$ and $\mathbf{G}$ with the vectors $|\tau(\ell-1)\rangle$ and $|\eta(\ell-1)\rangle$ (Fig. 1) to obtain $|\tau(\ell)\rangle$ and $|\eta(\ell)\rangle$, so we use this step to estimate the time complexity of the algorithm. Assume that each state has an average of γ outgoing jumps, so that the jump matrix $\mathbf{Q}$ has γN nonzero entries. Each transfer matrix has $(n_{\max} + 1)(n_{\max} + 2)/2$ nonzero blocks (Eq. (43)), yielding approximately $\gamma N(n_{\max} + 1)(n_{\max} + 2)/2$ total nonzero entries in $\mathbf{K}$ and $\mathbf{G}$. Since we multiply these matrices by the $N(n_{\max} + 1)$-dimensional state vectors at each of the Λ total jumps, the algorithm scales as

$$O\left(n_{\max}^2 \gamma N \Lambda\right). \qquad (59)$$

Assuming there are loops in the model (i.e., there is no maximum possible path length), the cutoff Λ scales linearly with the mean path length $\bar{\ell}$, as well as the maximum moment $n_{\max}$ as argued in Sec. III C. For simple random walks, the mean path length scales as a power of the total number of states:

$$\bar{\ell} \sim \begin{cases} N^{d_{\mathrm{w}}/d_{\mathrm{f}}} & \text{for } d_{\mathrm{w}} > d_{\mathrm{f}}, \\ N & \text{for } d_{\mathrm{w}} \leq d_{\mathrm{f}}, \end{cases} \qquad (60)$$

where d_{w} is the dimension of the random walk and d_{f} is the fractal dimension of the space.[11,14] Strictly speaking, these scaling relations depend on the boundary conditions (proximity of the initial and final states) and the presence of an energy landscape; the scaling relations in Eq. (60) are a "worst-case scenario" when the landscape is flat and the initial and final states are very far from each other. Altogether this

implies that the algorithm will scale as

$$
\begin{cases}
O(n_{\max}^3 \gamma N^{1+d_w/d_f}) & \text{for } d_w > d_f, \\
O(n_{\max}^3 \gamma N^2) & \text{for } d_w \le d_f.
\end{cases} \tag{61}
$$

Alternative recursive expressions for first-passage time moments on 1D lattices scale as $O(N^3)$,[41] as do general methods for solving the backward equation (a linear system) in transition path theory;[42] our scaling will be equivalent in the extreme case of a fully connected network where $\gamma = N - 1$.

IV. EXAMPLES

We now illustrate the path ensemble approach on a series of simple examples.

A. 1D lattice

We first consider a Markov CTRW on a 1D lattice. Let the lattice have L sites with equal and symmetric transition rates between neighboring sites:

$$
\begin{aligned}
\langle x+1 | \mathbf{W} | x \rangle &= 1 & \text{for } 1 \le x < L, \\
\langle x-1 | \mathbf{W} | x \rangle &= 1 & \text{for } 1 < x \le L, \\
\langle y | \mathbf{W} | x \rangle &= 0 & \text{otherwise.}
\end{aligned} \tag{62}
$$

From the rate matrix $\mathbf{W}$, we can obtain the jump matrix $\mathbf{Q}$ and the waiting time moments $\theta^{(n)}$ using Eqs. (2) and (5); note that the reflecting boundary conditions mean the "bulk" states ($1 < x < L$) have $\theta_{\text{bulk}}^{(1)} = 1/2$, while the "edge" states

($x = 1$, $x = L$) have $\theta_{\text{edge}}^{(1)} = 1$ due to their different connectivities (numbers of outgoing jumps). We consider the ensemble of first-passage paths from one end of the lattice ($x = 1$) to the other ($x = L$). Figure 2(a) shows the distributions $\bar{t}^{(n)}(\ell)$ of path time moments over path lengths; the path length probability distribution $\rho(\ell) = \bar{t}^{(0)}(\ell)$ is very close to exponential except for small ℓ, while the higher moments illustrate the Erlang-like function derived in Eq. (58). In particular, we confirm that the higher time moments decay approximately exponentially for large ℓ. Since the connectivity is nearly the same everywhere for large L ($\gamma = 2$ for all states except $x = 1$ and $x = L$), Eq. (30) indicates the distribution of path action will also be exponential with mean action (path entropy) $\approx \bar{\ell} \log 2$.

We now introduce a potential energy $V(x) = (L - x)/(L - 1)$ that provides a constant force down the lattice (Fig. 2(b), inset). If we use Metropolis transition rates[38] $\langle y | \mathbf{W} | x \rangle = \min(1, e^{-\beta(V(y)-V(x))})$, where β is the inverse temperature, we obtain a biased random walk with forward rate of 1 and backward rate of $e^{-\beta/(L-1)}$:

$$
\begin{aligned}
\langle x+1 | \mathbf{W} | x \rangle &= 1 & \text{for } 1 \le x < L, \\
\langle x-1 | \mathbf{W} | x \rangle &= e^{-\beta/(L-1)} & \text{for } 1 < x \le L, \\
\langle y | \mathbf{W} | x \rangle &= 0 & \text{otherwise.}
\end{aligned} \tag{63}
$$

As we increase the inverse temperature β from zero, the bias becomes exponentially stronger, leading to a distribution of path lengths more tightly concentrated around the minimum length $\ell = L - 1$ (Fig. 2(b)). Since only a single path with probability 1 is available in the limit $\beta \to \infty$, the distributions
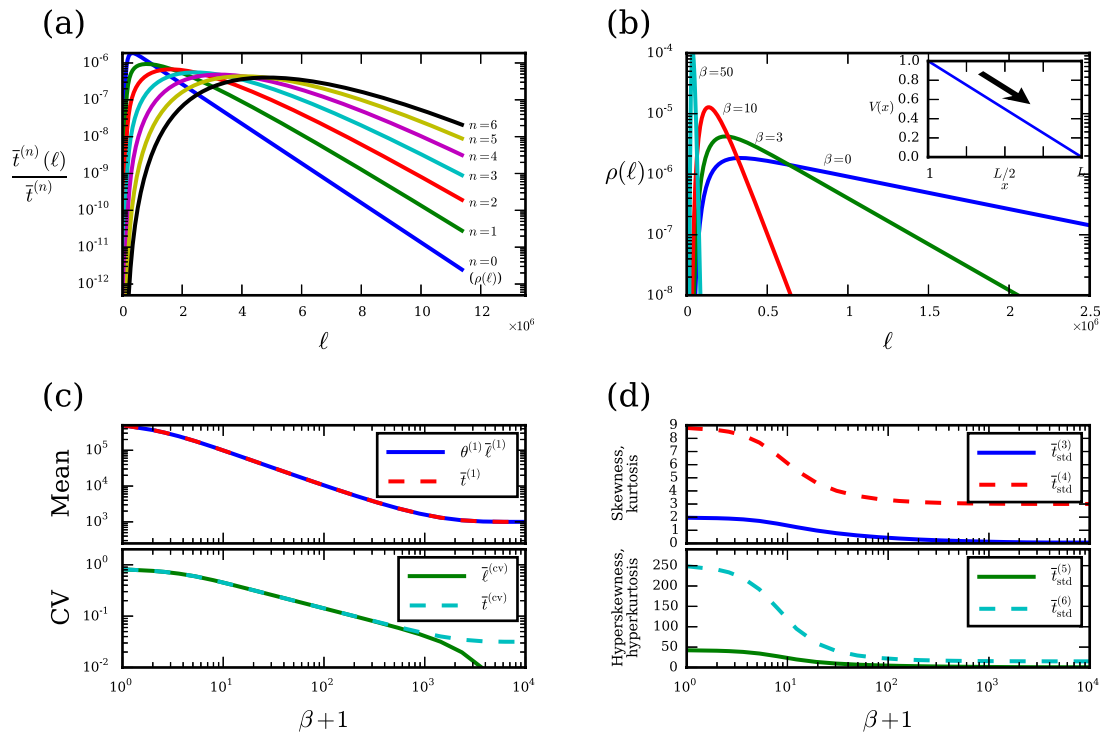

FIG. 2. Distributions of path lengths and times on a 1D lattice. (a) The nth time moments $\bar{t}^{(n)}(\ell)$ for paths of length ℓ, normalized as fractions of the total moments $\bar{t}^{(n)}$, in the absence of a potential energy. (b) Path length probability distribution $\rho(\ell)$ for several choices of β on a linear energy landscape $V(x)$ (inset). (c) The mean path length $\bar{\ell}^{(1)}$ (scaled by the mean waiting time $\theta^{(1)} = 1/2$ for bulk states), mean path time $\bar{t}^{(1)}$, path length CV $\bar{\ell}^{(cv)}$, and path time CV $\bar{t}^{(cv)}$ as functions of β. (d) Skewness $\bar{t}_{\text{std}}^{(3)}$, kurtosis $\bar{t}_{\text{std}}^{(4)}$, hyperskewness $\bar{t}_{\text{std}}^{(5)}$, and hyperkurtosis $\bar{t}_{\text{std}}^{(6)}$ of path time as functions of β. Values of β in (c) and (d) are shifted by 1 to show $\beta = 0$ on a log scale. In all panels, we use a lattice of length $L = 1000$ with transition rates given by Eq. (63).

of path lengths and actions become delta functions; in particular, path entropy is zero because the process is completely deterministic.

What is the distribution of path times as a function of β? In Fig. 2(c), we show the mean path time $\bar{t}^{(1)}$, which decreases dramatically as the bias is increased through β. It is almost exactly proportional to the mean path length $\bar{\ell}^{(1)}$ for all β, as predicted by Eq. (26) since $\theta^{(cv)} = 1$ and $\bar{\ell}^{(1)} = \bar{\ell} \gg 1$. We also show the CVs $\bar{\ell}^{(cv)}$, $\bar{t}^{(cv)}$ of the path length and time distributions, which measure the dispersion. For $\beta = 0$ both CVs are very close to 1, suggesting that the distributions of path lengths and times are approximately exponential. However, as β becomes large, the length CV $\bar{\ell}^{(cv)}$ drops to zero, since the length distribution becomes a delta function, but the time CV $\bar{t}^{(cv)}$ decreases to a small but nonzero value, indicating that the distribution becomes narrowly but finitely distributed around its mean.

Besides CV, standardized moments offer a useful way to characterize the shape of a distribution.[43] They are defined as dimensionless moments of a random variable X shifted and rescaled to have mean 0 and standard deviation 1:

$$\langle X^n \rangle_{\text{std}} = \frac{\langle (X - \langle X \rangle)^n \rangle}{\left[\left\langle (X - \langle X \rangle)^2 \right\rangle \right]^{n/2}}. \tag{64}$$

Since the first and second standardized moments are 0 and 1 by construction, the lowest non-trivial moment is the third moment, traditionally known as skewness since it measures the asymmetry of the distribution around the mean. The fourth standardized moment is the kurtosis; the fifth and sixth standardized moments are sometimes called the hyperskewness and hyperkurtosis. For an exponential distribution, the nth standardized moment is $!n$, i.e., the subfactorial or the number of derangements of n objects. Therefore, exponential skewness, kurtosis, hyperskewness, and hyperkurtosis are 2, 9, 44, and 265, respectively. For a Gaussian distribution, the first four standardized moments are 0, 3, 0, and 15, respectively.

Figure 2(d) shows the first four non-trivial $(n \geq 3)$ standardized moments of path time on the 1D lattice as functions of β. For $\beta = 0$ the standardized moments are very close to their exponential values, confirming that the distribution of first-passage times for a simple random walk is very close to exponential. However, as we increase the rightward bias by increasing β, the moments undergo a rapid transition near $\beta \approx 10$. Note that this transition happens at rather low temperature $(T = \beta^{-1} = 10^{-1})$ compared to the total change in energy across the lattice, which is set to 1. For very large β, the standardized time moments saturate at the Gaussian values of 0, 3, 0, and 15. This is because a single path of minimal length $\ell = L - 1$ dominates at low temperatures (Fig. 2(b)), and thus the total path time is just the sum of the waiting times along the single path. By the central limit theorem, this sum will be approximately Gaussian for large L. Since $\theta_{\text{edge}}^{(1)} = \theta_{\text{bulk}}^{(1)} = \theta_{\text{edge,c}}^{(2)} = \theta_{\text{bulk,c}}^{(2)} = 1$ in this limit (bulk and edge states are the same since travel along the lattice is one-way), the mean and variance of path time in this limit should be the path length $\ell = L - 1 = 999$ times the mean and variance of each waiting time: $\bar{t}^{(1)} = (L - 1)\theta_{\text{bulk}}^{(1)} = 999$

and $\bar{t}_c^{(2)} = (L - 1)\theta_{\text{bulk,c}}^{(2)} = 999$, leading to a coefficient of variation $\bar{t}^{(cv)} = \sqrt{999}/999 \approx 0.03$. This agrees with Fig. 2(c). This simple model is reminiscent of downhill folding in proteins[17] and linear biochemical pathways such as those used in kinetic proofreading,[44] where non-exponential kinetics and the transition between exponential and deterministic (narrow Gaussian distribution) regimes have been previously investigated.

Equation (26) suggests that the length and time distributions should be very similar even for $\beta \to \infty$, since the correction term is still small ($\theta^{(cv)} = 1$ and $\bar{\ell} = L - 1 = 999 \gg 1$). Indeed, the time distribution is a Gaussian narrowly distributed around its mean, whereas the length distribution is a delta function; the differences in the moments are of the order $1/\bar{\ell} \approx L^{-1}$. However, this slight difference is better resolved by considering the complete relation between length and time moments (Eq. (24)) with cumulants $\langle \mathcal{T}^{(n)} \rangle_c$ and $\langle \mathcal{L}^n \rangle_c$ instead of the raw moments. For $\beta \to \infty$, all $\langle \mathcal{L}^n \rangle_c = 0$ for $n \geq 2$, which means that we cannot expand Eq. (24) for the cumulants as in Eq. (26). Instead, the only nonzero terms in Eq. (24) yield the exact equation $\langle \mathcal{T}^{(n)} \rangle_c = \theta_c^{(n)} \langle \mathcal{L} \rangle_c$ for all n.

B. 1D comb and memory from coarse-graining

We now turn to an example that explores the effects of waiting memory on distributions of path times. We consider the 1D comb: a 1D backbone of length L_{backbone} where each site has a 1D tooth of length L_{tooth} extending from it (Fig. 3(a)). Combs have traditionally represented simple models of diffusion on percolation clusters and other fractal structures in disordered materials;[4,45] more recently, they have also been proposed as a model for cancer cell proliferation.[46] As in the previous example (Eq. (62)), we use symmetric transition rates of 1 between neighboring sites in the comb. If we are primarily interested in diffusion along the backbone rather than within the teeth, it is natural to coarse-grain each tooth into a single effective backbone state with some effective waiting time distribution $\psi(t)$ that describes the time spent exploring the tooth before returning to make a jump along the backbone (Fig. 3(a)).[4,45] The waiting times within each coarse-grained backbone state are therefore the first-passage times to return to the backbone after exploring the tooth. The distribution of these return times, $f_{\text{tooth}}(t)$, has the approximate form[4]

$$\psi(t) = f_{\text{tooth}}(t) \sim \begin{cases} t^{-3/2} & \text{for } t < \tau L_{\text{tooth}}^2, \\ e^{-t/(\tau L_{\text{tooth}}^2)} & \text{for } t > \tau L_{\text{tooth}}^2, \end{cases} \tag{65}$$

where τ is a time scale that is $O(1)$ in L_{tooth}. Since the distributions of path times and lengths are very similar on 1D lattices in the absence of potential (cf. the $\beta = 0$ limit in Fig. 2(c)), the crossover time τL_{tooth}^2 is essentially the characteristic time scale to explore a 1D lattice of length L_{tooth} (Eq. (60)). In Fig. 3(b) we show the path length distribution $\rho_{\text{tooth}}(\ell)$ to exit the tooth, which according to Eq. (26) should be approximately the same as the distribution of times $f_{\text{tooth}}(t)$ for large L_{tooth}. Indeed, $\rho_{\text{tooth}}(\ell)$ follows the form of Eq. (65) very clearly: there is a power law regime of $\ell^{-3/2}$ until
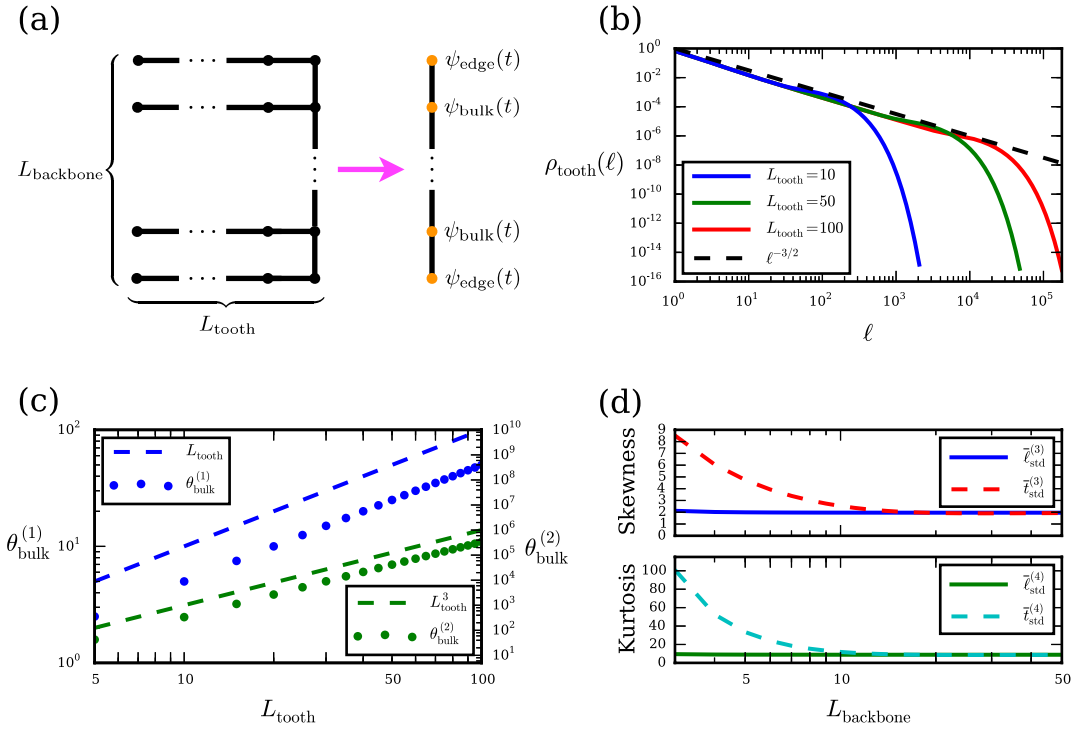
FIG. 3. Distributions of path lengths and times on a 1D comb. (a) Schematic of a comb with backbone of length L_{backbone} and teeth of length L_{tooth}. We coarse-grain the teeth into single states (orange) along the backbone with effective waiting time distributions $\psi_{\text{edge}}(t)$ and $\psi_{\text{bulk}}(t)$. (b) Path length distribution $\rho_{\text{tooth}}(\ell)$ to exit a tooth of different lengths L_{tooth}, along with the power law $\ell^{-3/2}$ for comparison. (c) Mean $\theta_{\text{bulk}}^{(1)}$ and second moment $\theta_{\text{bulk}}^{(2)}$ of $\psi_{\text{bulk}}(t)$ as functions of tooth length L_{tooth}; points are numerical calculations, while dashed lines show expected scaling behavior from Eq. (66). (d) Skewness $\bar{\ell}_{\text{std}}^{(3)}$, $\bar{t}_{\text{std}}^{(3)}$ and kurtosis $\bar{\ell}_{\text{std}}^{(4)}$, $\bar{t}_{\text{std}}^{(4)}$ of length and time distributions for paths along the backbone as functions of L_{backbone}, with $L_{\text{tooth}} = 100$.

approximately $\ell \sim L_{\text{tooth}}^2$, after which there is an exponential decay. To estimate the tooth waiting time moments $\theta^{(n)}$ from $\psi(t) = f_{\text{tooth}}(t)$, we make the approximation that the power-law regime dominates the moment integrals:[4]

$$\theta^{(n)} \sim \int^{\tau L_{\text{tooth}}^2} dt \, t^{-3/2} t^n \sim L_{\text{tooth}}^{2n-1}. \tag{66}$$

In Fig. 3(c) we verify this scaling by numerically calculating the moments from first-passage paths that exit the tooth.

The dominant power-law regime of $\psi(t)$ means that its statistics are very different from those of an exponential distribution. For example, the CV is $\theta^{(\text{cv})} \sim L_{\text{tooth}}^{1/2}$ rather than ~ 1, indicating a much broader distribution of times compared to the exponential case. The non-exponential nature of the waiting time distribution is indicative of memory within an effective backbone state: how much longer the system waits in the state depends on how long it has already waited. Mathematically, this apparent memory arises from coarse-graining each tooth into a single state, which erases information about the position of the system within the tooth. Indeed, for the distribution of times in Eq. (65), the mean waiting time starting from $t = 0$ is $\sim L_{\text{tooth}}$, since it is dominated by the power-law regime (Eq. (66)). However, if the system waits at least time $\sim L_{\text{tooth}}^2$, the mean additional waiting time becomes $\sim L_{\text{tooth}}^2$, due to the exponential regime. In other words, if the system does not leave by time $\sim L_{\text{tooth}}^2$—meaning that it has diffused far from the backbone—it is likely to wait much longer as the exponential regime of $\psi(t)$ takes over.

One effect of this memory is that it can lead to significant differences between the distributions of path times and path lengths along the effective backbone states. Equation (26) shows that the moments of path time and path length are approximately proportional if the waiting time distributions are not too broad relative to ratios of path length moments; that is, the correction term in Eq. (26) is small if $(\theta^{(\text{cv})})^2 \ll \langle \mathcal{L}^n \rangle / \langle \mathcal{L}^{n-1} \rangle$. We estimate the size of this correction for the comb model, focusing on first-passage paths from one end of the backbone to the other. The waiting time CV is $\theta^{(\text{cv})} \sim L_{\text{tooth}}^{1/2}$ as previously mentioned. The path length distribution, meanwhile, appears to be very close to exponential: Fig. 3(d) shows that its skewness and kurtosis are consistent with their exponential values (2 and 9) for any backbone length. Since the mean path length should be $\langle \mathcal{L} \rangle = \bar{\ell} \sim L_{\text{backbone}}^2$ (Eq. (60)), this implies that the higher moments are $\langle \mathcal{L}^n \rangle \sim n! L_{\text{backbone}}^{2n}$. Therefore, the correction term in Eq. (26) is approximately

$$\binom{n}{2} (\theta^{(\text{cv})})^2 \frac{\langle \mathcal{L}^{n-1} \rangle}{\langle \mathcal{L}^n \rangle} \sim \frac{1}{2}(n-1) \frac{L_{\text{tooth}}}{L_{\text{backbone}}^2}. \tag{67}$$

Thus, when $L_{\text{tooth}} \ll L_{\text{backbone}}^2$, we expect path lengths and times along the backbone to have similar statistics, with a pronounced difference in the opposite limit. In Fig. 3(d), we calculate skewness and kurtosis of path time moments, varying L_{backbone} while fixing $L_{\text{tooth}} = 100$. Indeed, for $L_{\text{backbone}} < \sqrt{L_{\text{tooth}}} = 10$, there is a large discrepancy between path length and time moments, while for $L_{\text{backbone}} > 10$, they become very close. The fact that the length of the teeth must

be large compared to the *square* of the backbone length to have an appreciable effect on the path statistics indicates that dynamics along the backbone, rather than within the teeth, tend to dominate the first-passage process.

C. Memory in coarse-grained metastable states

As the previous example showed, memory, in the form of non-exponential distributions of waiting times, naturally arises from coarse-graining because information about the microscopic states of the system is lost. This principle plays a crucial role in the generation of discrete stochastic models for protein folding and other molecular processes.[16] In these models, a high-dimensional space of "microscopic" states (e.g., protein conformations) is coarse-grained into a discrete set of "macroscopic" states with some effective transition probabilities. The resulting coarse-grained model is more amenable to calculating statistical properties of dynamics over long time scales, such as the mean protein folding time or kinetic bottlenecks.[1,2] However, the coarse-graining can result in qualitative differences between the approximate macroscopic model and the true underlying microscopic dynamics,[16] including non-exponential waiting times in the effective states that are not addressed by conventional transition path theory.[15]

As a simple illustration of this phenomenon, we consider a 2D double-well potential

$$V(x, y) = (x^2 - 1)^2 + 2y^2, \qquad (68)$$

as shown in Fig. 4(a). This potential has two local minima at $(\pm 1, 0)$ with $V = 0$, a central barrier at $(0,0)$ with $V = 1$, and reflecting boundaries at $x, y = \pm 2$. At low temperatures, the system will spend most of its time in the basins around the two minima. Therefore, it is natural to coarse-grain the "microscopic" 2D space into two metastable states, A and B, separated by the central energy barrier (Fig. 4(a)). To characterize the statistics of the two-state dynamics, it is common to calculate a single reaction rate (inverse of the mean time) from one state to another. However, using a single rate parameter implicitly assumes that waiting times in the coarse-grained states are distributed exponentially. Here we show this to be a poor approximation.

When the system first transitions to B from A, it starts just to the right of the interface separating the two states (Fig. 4(a)). The effective waiting time in the coarse-grained state B is therefore the time until the system first returns to that interface, starting from one step off it. We explicitly calculate the first-passage paths for this microscopic process using our numerical method. We discretize the space into a 2D lattice with $\Delta x = 0.05$ and assume a Markov CTRW on the lattice with Metropolis transition rates for jumps between nearest neighbors. Although the system can enter state B at any point along the interface with A, for simplicity we assume that it entered through the central barrier at $(0,0)$ (the lowest-energy point along the boundary) and therefore starts in B at $(\Delta x, 0)$ (marked by the blue square in Fig. 4(a)). In Fig. 4(b) we show the distributions of path lengths (blue line) starting from this point and returning anywhere along the interface for low
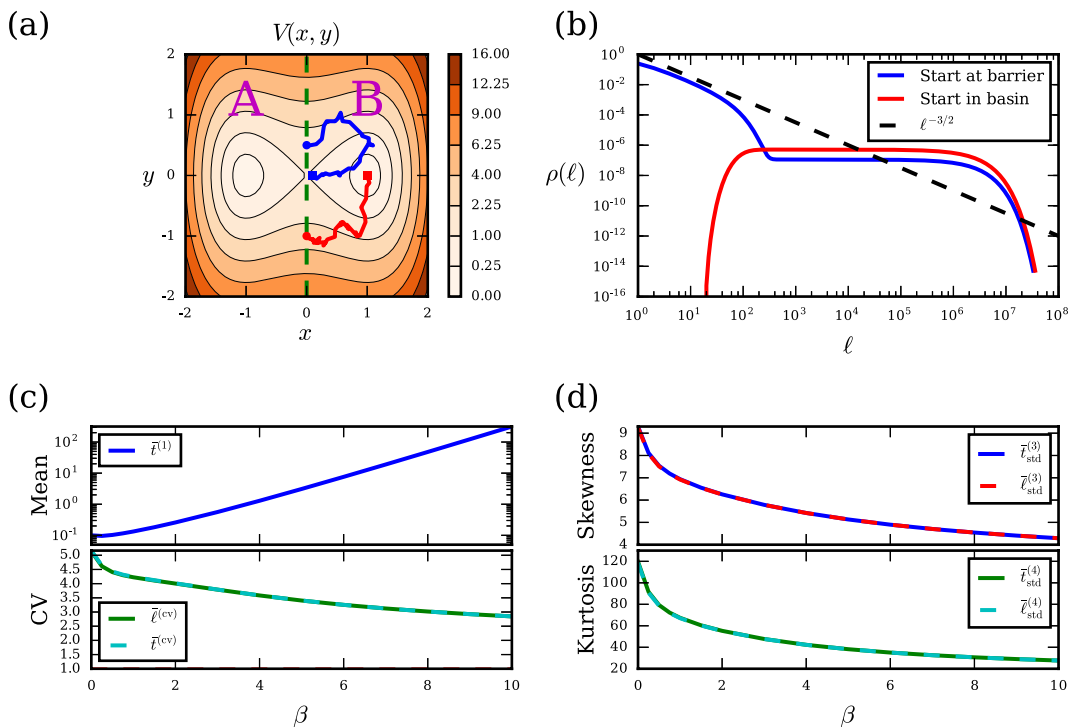

FIG. 4. Effect of memory in coarse-grained metastable states. (a) Double-well potential (Eq. (68)) exactly coarse-grained into states A and B with boundary along the green dashed line. We also show example paths (solid blue and red lines) that both exit B (circles) but start from different initial conditions within B (squares). (b) Length distributions $\rho(\ell)$ of paths (with $\beta = 10$) that exit B but start at different initial conditions (at the central barrier, corresponding to the blue square in (a), or in the low-energy basin, corresponding to the red square in (a)), along with the power law $\ell^{-3/2}$ for comparison. For paths that exit the coarse-grained state B, (c) the mean time $\bar{t}^{(1)}$ and CVs $\bar{\ell}^{(\text{cv})}$, $\bar{t}^{(\text{cv})}$ of length and time, and (d) skewness $\bar{\ell}_{\text{std}}^{(3)}$, $\bar{t}_{\text{std}}^{(3)}$ and kurtosis $\bar{\ell}_{\text{std}}^{(4)}$, $\bar{t}_{\text{std}}^{(4)}$ of length and time, all as functions of β. All calculations use a discretized square lattice with $\Delta x = 0.05$ over the space $(x, y) \in [-2, 2] \times [-2, 2]$.

temperature ($\beta = 10$); an example of such a path is shown in Fig. 4(a) (blue line). The distribution has a power-law regime for small ℓ and exponential regime for large ℓ. These two asymptotic limits are the same as the distribution of waiting times in the 1D tooth (Eq. (65), Fig. 3(b)), and indeed, they have a similar physical basis: the power-law regime arises from paths that quickly return to the interface without falling into the low-energy basin, while the exponential regime arises from paths that fall into the low-energy basin before returning. In contrast to the 1D comb, though, there is a broad flat region of the distribution between the power-law and exponential regimes. This is actually part of the distribution of paths that fell into the low-energy basin before returning; it corresponds to an intermediate regime before that distribution hits its asymptotic exponential tail (cf. path length distributions on a 1D lattice in Fig. 2(b)). We confirm this by directly calculating paths with the starting point in the basin (red line in Fig. 4(b); an example path is shown in Fig. 4(a) from the red square to the red circle).

In Figs. 4(c) and 4(d), we demonstrate that the distribution of path times, i.e., the waiting times in the coarse-grained state B (or A by symmetry), is nearly identical to the path length distribution shown in Fig. 4(b). Besides the mean time $\bar{t}^{(1)}$, we also calculate the CV and standardized moments for both length and time distributions, which are indistinguishable over the entire range of β (despite the heterogeneities in waiting time distributions across the lattice). Hence, the path length distribution in Fig. 4(b) also describes the effective waiting time distribution in the coarse-grained state. The power-law regime of this distribution for short paths leaves a distinct signature in the moments. Even at $\beta = 10$, which represents a temperature that is 10 times smaller than the lowest energy barrier (such that we expect the metastable approximation of A and B to be very good), the distribution of times deviates strongly from an exponential distribution: the CV is nearly 3, while the skewness and kurtosis are much larger than their exponential expectations. At lower β (higher temperatures), the deviation from an exponential distribution becomes even more pronounced. As with the comb, this enrichment of the distribution for very short paths means that given the system just transitioned to B, it is likely to quickly transition back to A. But if it does not transition back quickly, it is likely to wait much longer as it falls into the basin and the waiting times become exponentially distributed.

From the comb and double-well examples, we can deduce some general principles for the waiting memory that results from coarse-graining a state space. Assume that the microscopic state space for a system is d-dimensional Euclidean space, which we coarse-grain into N_{macro} effective macroscopic states, each consisting of N_{micro} microscopic states. The interfaces between coarse-grained states have dimension $d - 1$. When the system first enters one of these coarse-grained states, it begins just inside an interface. Therefore, the waiting time distribution $\psi(t)$ in the coarse-grained state is the first-passage time to return to that $(d - 1)$-dimensional interface. This return process is effectively a 1D random walk, since only the direction normal to the interface matters (at least within a neighborhood of the initial state, assuming the interface is locally flat). Therefore, the

distribution of first-passage times to return to the interface will be the same as for the 1D tooth in the comb (Eq. (65)):

$$\psi(t) \sim \begin{cases} t^{-3/2} & \text{for } t < \tau N_{\text{micro}}^{\nu}, \\ e^{-t/(\tau N_{\text{micro}}^{\nu})} & \text{for } t > \tau N_{\text{micro}}^{\nu}, \end{cases} \qquad (69)$$

where the crossover time between these regimes is the characteristic time scale $\tau N_{\text{micro}}^{\nu}$ to explore the coarse-grained state (Eq. (60); $\nu = 2$ for $d = 1$, $\nu = 1$ for $d \geq 2$), and τ is a microscopic time scale which is $O(1)$ in N_{micro}. The waiting time moments are approximately

$$\theta^{(n)} \sim \int^{\tau N_{\text{micro}}^{\nu}} dt \, t^{-3/2} t^n \sim N_{\text{micro}}^{\nu(n-1/2)}. \qquad (70)$$

In particular, the CV is

$$\theta^{(\text{cv})} = \frac{\sqrt{\theta_{\text{c}}^{(2)}}}{\theta^{(1)}} \sim N_{\text{micro}}^{\nu/4} = \begin{cases} N_{\text{micro}}^{1/2} & \text{for } d = 1, \\ N_{\text{micro}}^{1/4} & \text{for } d \geq 2. \end{cases} \qquad (71)$$

As with the 1D comb, the CV scales as a power of the coarse-grained state size, but rather slowly.

We can now determine whether such waiting time distributions will lead to different statistics of path lengths and times in the coarse-grained model. Since the mean path length in the coarse-grained model is $\bar{\ell} \sim N_{\text{macro}}^{\nu}$ (assuming the microscopic and coarse-grained spaces have the same dimensionality), the condition $(\theta^{(\text{cv})})^2 \ll \bar{\ell}$ (Eq. (29)) for equivalent path length and time statistics becomes

$$N_{\text{micro}} \ll N_{\text{macro}}^2. \qquad (72)$$

This is consistent with the condition found for the 1D comb where $N_{\text{micro}} = L_{\text{tooth}}$ and $N_{\text{macro}} = L_{\text{backbone}}$. For the double-well model, $N_{\text{micro}} \approx 3200$ (number of microscopic lattice points in A or B) and $N_{\text{macro}} = 2$; Eq. (72) does not hold in this case, so we expect significant differences in the statistics of path lengths (jumps between A and B) and path times in the coarse-grained model. In general, Eq. (72) implies that the more coarse-graining there is (resulting in fewer but larger coarse-grained states), the more significant the memory effects are on the effective CTRW.

D. Random barrier model (RBM) and memory from spatial disorder

To further demonstrate the effects of a complex energy landscape on path statistics, we consider the random barrier model (RBM),[3,30] a simple model of transport in disordered systems. In this model, a particle diffuses across a lattice with quenched energy barriers of random height between neighboring points. Here we consider a 2D lattice with energy barriers drawn from an exponential distribution $p(E) = E_0^{-1} e^{-E/E_0}$, where E_0 is the average energy.[47] We assume a Markov CTRW on this lattice with symmetric transition rates between neighboring states that depend exponentially on the intervening energy barrier:

$$\langle x', y' | \mathbf{W} | x, y \rangle = \Gamma_0 e^{-\beta E(x', y'; x, y)}, \qquad (73)$$

where Γ_0 is the rate of traversing a barrier of zero height (maximum possible rate) and $E(x', y'; x, y)$ is the energy barrier between (x', y') and (x, y). We use reflecting boundary

conditions and set $\Gamma_0 = E_0 = 1$ without loss of generality, as these two quantities set the overall time and energy scales. From the rates in Eq. (73), we determine jump probabilities and exponential waiting time moments using Eqs. (2) and (5).

Figure 5 shows a single (quenched) realization of the RBM on a 10×10 lattice for different values of β. In each panel, cells correspond to lattice points, while the gray-scale bars between them indicate the height of the intervening energy barriers (same in all panels). Due to the exponential distribution of energies, most barriers are low (black), with only a few relatively high barriers (white). We consider the ensemble of first-passage paths on this landscape from $(1,1)$ (bottom-left corner) to $(10,10)$ (top-right corner). First, we determine path statistics for $\beta = 0$ (Fig. 5(a)), where all transition rates are equal and the barriers have no effect (Eq. (73)). The leftmost panel of Fig. 5(a) shows the mean waiting time $\theta^{(1)}(x,y)$ in each state. When all transition rates

are equal, $\theta^{(1)}(x,y)$ depends only on the state's connectivity: the states in the bulk with more neighbors have shorter mean waiting times than do the edge and especially the corner states, which have fewer neighbors. The middle panel of Fig. 5(a) shows the average number of visits $v(x,y)$ to each state during the first-passage process (Sec. III B). For $\beta = 0$, the number of visits depends on both the distance to the final state and the state's connectivity: edge and corner states with fewer neighbors are visited less often than are bulk states the same distance from the final state. When we consider the mean fraction of time spent in each state (the product of $\theta^{(1)}(x,y)$ and $v(x,y)$, normalized by the total mean path time $\bar{t}^{(1)}$; rightmost panel of Fig. 5(a)), the connectivity-dependence largely disappears, so that the fraction of time depends mostly just on the distance to the final state.

For $\beta > 0$, the effects of the random energy barriers emerge. In Figs. 5(b) and 5(c), we show path statistics for
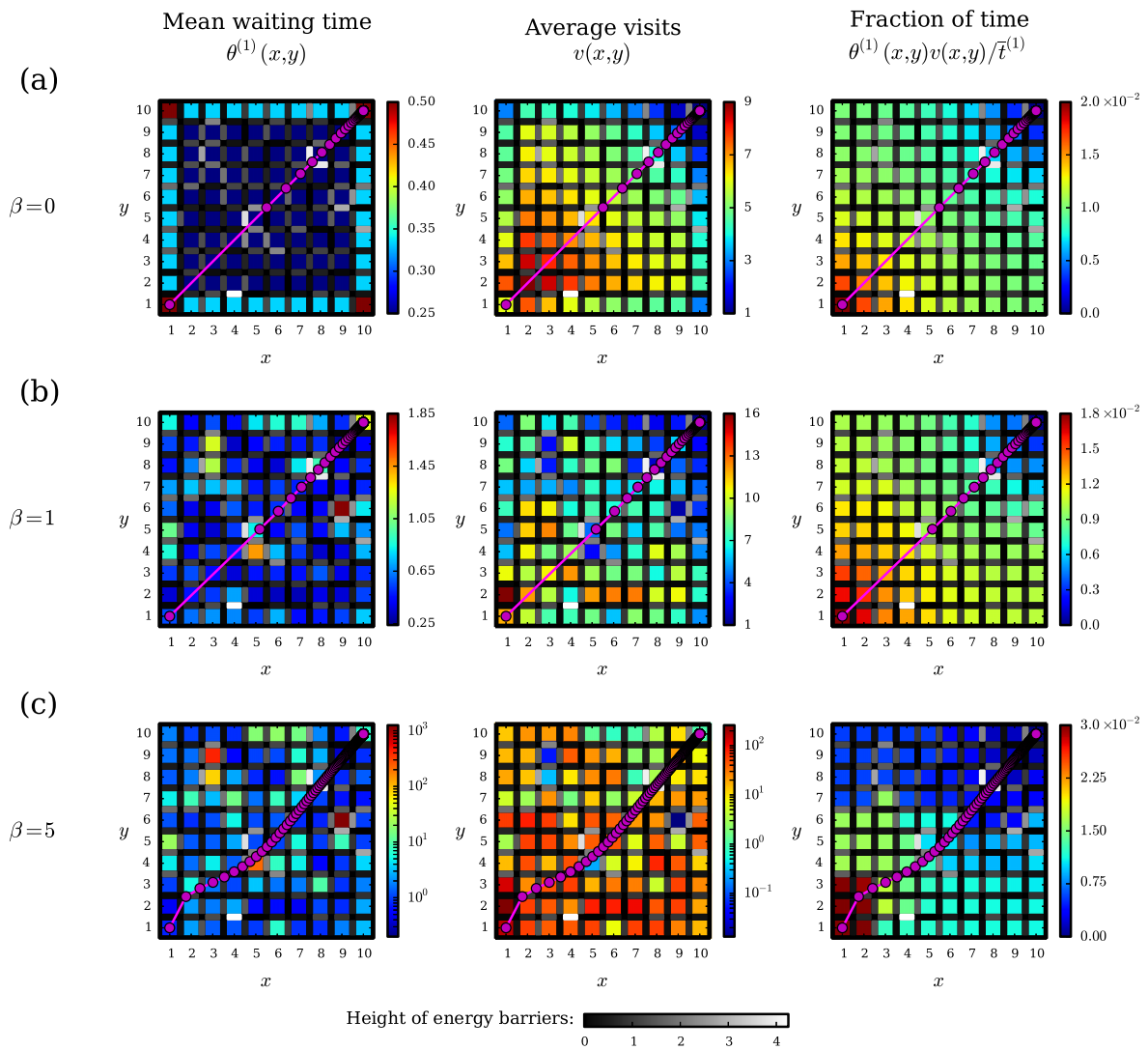

FIG. 5. Spatial properties of first-passage paths in the random barrier model. For a 10×10 lattice, we show statistics of first-passage paths from $(1,1)$ to $(10,10)$ for a single quenched realization of the energy barriers. Each colored cell corresponds to a lattice point (x,y), with gray-scale bars indicating energy barriers between lattice points (higher energies are white, lower energies are black). Energy barriers are randomly sampled from an exponential distribution with mean $E_0 = 1$, and the transition rate across a zero energy barrier is $\Gamma_0 = 1$. The leftmost column shows the mean waiting time $\theta^{(1)}(x,y)$, the middle column is the average number of visits $v(x,y)$, and the rightmost column is the average fraction of time $\theta^{(1)}(x,y)v(x,y)/\bar{t}^{(1)}$ spent at each lattice point. Rows correspond to different inverse temperatures: (a) $\beta = 0$, (b) $\beta = 1$, and (c) $\beta = 5$. Magenta points show the average particle position for every 100th jump (connected by straight lines to guide the eye).

$\beta = 1$ and $\beta = 5$. States with large barriers around them acquire significantly longer mean waiting times, leading to a very broad distribution of time scales; at $\beta = 5$ the mean waiting times span three orders of magnitude (Fig. 5(c), leftmost panel). However, states with extremely long mean waiting times also tend to have many fewer visits on average (Figs. 5(b) and 5(c), middle panels). This is because the high energy barriers that make these states difficult to exit also make them difficult to enter in the first place. In contrast with $\beta = 0$, where $v(x, y)$ was determined by both the distance to the final state and the state's local connectivity, for large β the average number of visits becomes predominately determined by the state's local properties, i.e., its mean waiting time, rather than its global position on the lattice. However, the heterogeneity of $\theta^{(1)}(x, y)$ and $v(x, y)$ across the lattice nearly vanishes when considering their product, the mean fraction of time (Figs. 5(b) and 5(c), rightmost panels): as with $\beta = 0$, the distance to the final state primarily determines the fraction of time spent at a lattice point. Instead of varying smoothly across states as for $\beta = 0$, though, at $\beta = 5$ the fraction of time appears to have four distinct plateaus on the 2D lattice. Within each plateau, the particle spends approximately the same fraction of total time at each lattice point.

Figure 5 also shows the average of all first-passage paths (magenta lines). We calculate the average path by defining state functions for each spatial coordinate as $B_x(x, y) = x$ and $B_y(x, y) = y$ and using Eq. (55) to determine the mean positions $\bar{B}_x(\ell)$ and $\bar{B}_y(\ell)$ as functions of the intermediate jump ℓ along a path. We plot these mean positions for every 100th jump in Fig. 5 to represent the average path of the particle. For $\beta = 0$, the average path is necessarily symmetric

across the diagonal and asymptotically converges toward the final state (Fig. 5(a)). For $\beta > 0$, the energy barriers slightly distort the average path at the beginning, but as the path approaches the final state, these asymmetries largely average out (Figs. 5(b) and 5(c)).

We next consider the distributions of path length, time, and action for the RBM. For $\beta = 0$, all of these distributions are close to exponential in shape, as expected from previous examples. For example, the length distribution has CV $\bar{\ell}^{(cv)} \approx 0.89$, skewness $\bar{\ell}^{(3)}_{std} \approx 1.99$, and kurtosis $\bar{\ell}^{(4)}_{std} \approx 8.95$. The moments for path time and action are also very close to these values: indeed, Eqs. (26) and (30) imply that these distributions should all be very similar since the network is nearly homogeneous.

What happens to these distributions in the presence of a complex energy landscape ($\beta > 0$)? Figure 6 shows distributions of the first four moments over many quenched realizations of the RBM at different β. For $\beta = 1$, both the mean path length and time are mostly close to their values at $\beta = 0$ (Fig. 6(a)), and their CVs and standardized moments indicate that the distributions are still close to exponential (Figs. 6(b)–6(d)). Larger β, however, leads to a very wide range of possible length and time moments, which can span several orders of magnitude across realizations. The correlations between path lengths and times are also significant. Mean lengths and times are mostly clustered along the diagonal, indicating their proportionality for most realizations, but there are some realizations with mean time much larger than mean length (Fig. 6(a)). For CVs and standardized moments, many realizations that deviate from exponential distributions do so equally in both length and time,
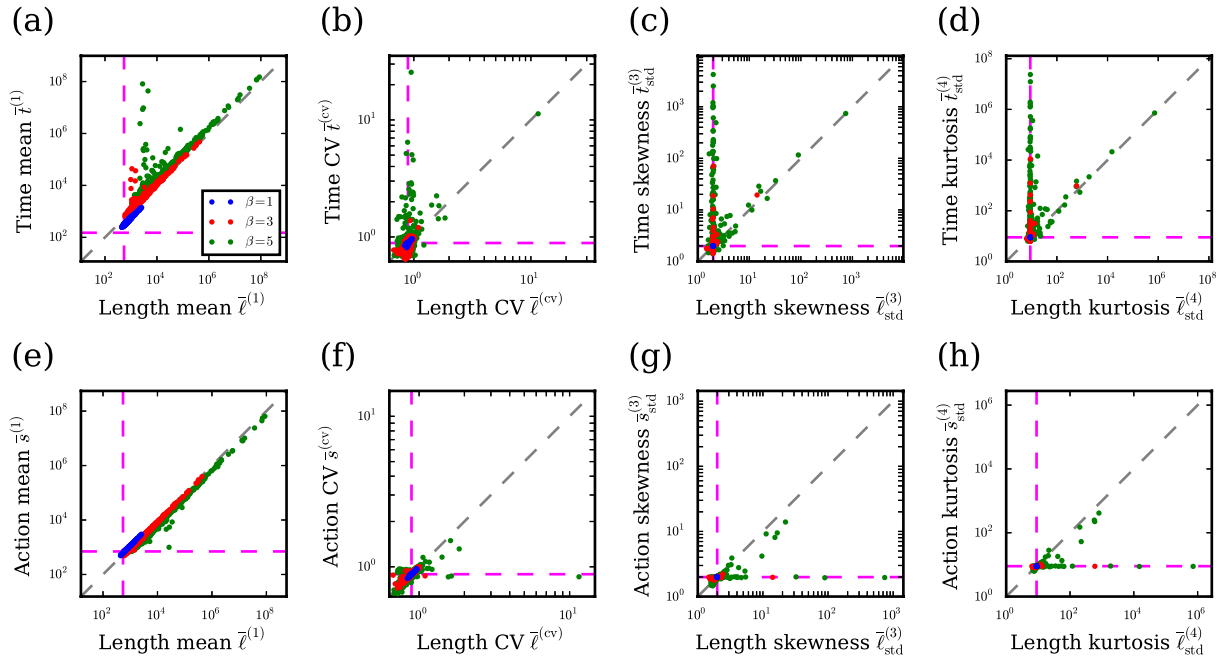

FIG. 6. Distributions of path statistics in the random barrier model. For 1000 quenched realizations of the energy barriers on a 10×10 lattice, we show: (a) mean path length $\bar{\ell}^{(1)}$ versus mean path time $\bar{t}^{(1)}$; (b) length CV $\bar{\ell}^{(cv)}$ versus time CV $\bar{t}^{(cv)}$; (c) length skewness $\bar{\ell}^{(3)}_{std}$ versus time skewness $\bar{t}^{(3)}_{std}$; (d) length kurtosis $\bar{\ell}^{(4)}_{std}$ versus time kurtosis $\bar{t}^{(4)}_{std}$; (e) mean length $\bar{\ell}^{(1)}$ versus mean action $\bar{s}^{(1)}$; (f) length CV $\bar{\ell}^{(cv)}$ versus action CV $\bar{s}^{(cv)}$; (g) length skewness $\bar{\ell}^{(3)}_{std}$ versus action skewness $\bar{s}^{(3)}_{std}$; and (h) length kurtosis $\bar{\ell}^{(4)}_{std}$ versus action kurtosis $\bar{s}^{(4)}_{std}$. Blue points are $\beta = 1$, red points are $\beta = 3$, and green points are $\beta = 5$; the horizontal and vertical dashed magenta lines correspond to the values of the moments for $\beta = 0$. We also show a diagonal gray line with slope 1 to guide the eye.

resulting in points along diagonal. However, there are also many realizations with highly non-exponential distributions of path times, even though the length distribution is close to exponential (Figs. 6(b)–6(d)). In this case, the approximate equivalence between path length and time in Eq. (26) breaks down not because $(\theta^{(cv)})^2 \gg \bar{\ell}$—this is never true in our RBM model since $\theta^{(cv)}(x, y) = 1$ for all (x, y) and $\bar{\ell}$ is always large—but because of the spatial disorder. Equation (26) is derived for a network with identical waiting time distributions at all states, but the RBM has a very broad range of mean waiting times for large β, as the example in Fig. 5(c) shows. Thus, a rugged energy landscape even with Markovian waiting times can lead to non-exponential path statistics, and hence the appearance of memory; such non-exponential kinetics have long been discussed in the context of glasses.[21,22]

Figures 6(e)–6(h) show similar distributions of moments for path action (plotted against path length moments for reference). Equation (30) shows that the moments of length and action are proportional for networks with homogeneous connectivity. While the 2D lattice in the RBM is not exactly homogeneous due to boundary conditions, Fig. 6(e) shows mean length and action to be very nearly proportional for almost all realizations. Since mean action is equivalent to path entropy, its wide range of possible values indicates that first-passage in some realizations is dominated by a few relatively high-probability paths, while in other realizations, it is dominated by a large number of much lower-probability paths. The higher moments of action in Figs. 6(f)–6(h) indicate that usually action is nearly exponentially distributed even for larger β. In particular, many of the realizations with non-exponential length distributions still have exponentially distributed actions. This suggests that the distribution of path actions is much more weakly affected by the energy landscape compared to path lengths and times.

V. DISCUSSION

We have studied CTRWs on networks using statistical mechanics of the path ensemble. A particular convenience of the path formalism lies in exploring the relationship between the distributions of path lengths and path times, which can be viewed as the relationship between the full continuous-time process and its discrete-time projection. Discrete-time models have generally dominated the theory of random walks not only due to their simplicity but also because we expect a continuous-time process on the same network to be nearly equivalent under certain conditions.[3,13] A well-known exception to this expectation is for waiting time distributions $\psi(t)$ without a characteristic time scale (divergent mean), which can produce anomalous diffusion even on regular lattices.[3] Using our approach, we have identified two more important exceptions. If all states have identical waiting time distributions $\psi(t)$, Eq. (26) shows that continuous- and discrete-time dynamics will have different statistics if $(\theta^{(cv)})^2 \gg \langle \mathcal{L}^n \rangle / \langle \mathcal{L}^{n-1} \rangle \sim \bar{\ell}$, where $\theta^{(cv)}$ is the CV of the waiting time distribution $\psi(t)$ and $\bar{\ell}$ is the mean path length. We should therefore expect significant differences between continuous- and discrete-time dynamics to occur

when $\psi(t)$ is much more broadly dispersed than an exponential distribution (large $\theta^{(cv)}$) and for small state spaces, which tend to produce small $\bar{\ell}$ (Eq. (60)). Furthermore, the 2D double-well example suggests this condition is still valid (Fig. 4(d)) even if the waiting time distributions and jump probabilities vary across states, as long as they do not vary too much. If they do, however, we find another exception to the equivalence of continuous- and discrete-time dynamics: even with exponential waiting times, spatial disorder can produce very different distributions of path lengths and path times, as illustrated in the random barrier model (Fig. 6).

Although we have focused primarily on moments of path statistics in this work, ideally we would like to know the entire distributions of these quantities. In principle, one can fit a parameterized distribution to the moments. In most statistical applications, this "method of moments" typically produces a good approximation for well-behaved distributions, especially using a very general parameterization such as the Pearson distribution.[43] For path distributions, a linear combination of exponential functions may be a more appropriate choice. Since path length, time, and action distributions are frequently very similar and since our method explicitly calculates the entire path length distribution already, fitting distributions from moments would be most valuable in cases where the continuous- and discrete-time processes are very different.

In any case, the moments of path statistics themselves are valuable for quantifying deviations from a simple exponential distribution. These deviations are important because they represent a form of memory: the amount of time for a process to occur depends on how much time has already passed. We have emphasized how coarse-graining many "microscopic" states of a system into a smaller number of effective "macroscopic" states generally leads to non-exponential $\psi(t)$ in the coarse-grained states; we explicitly demonstrated this by coarse-graining teeth in a 1D comb (Fig. 3) and low-energy basins in a double-well potential (Fig. 4). Furthermore, we have argued that $\psi(t)$ in coarse-grained states will frequently obey Eq. (69), with a power-law regime for short times and an exponential regime for long times. Physically, this distribution arises because the system always starts just inside the boundary of a coarse-grained state; therefore, it can either quickly recross the boundary, leading to the power-law regime, or explore the rest of the coarse-grained state, leading to the exponential regime. Compared to a simple exponential distribution, this hybrid distribution is enriched by the power law at short times, meaning that very short waiting times are much more likely than would be expected if the system started in the middle of the coarse-grained state rather than near the boundary. However, if the system does not quickly exit, it is likely to wait much longer as it explores the rest of the coarse-grained state. This effective $\psi(t)$ is typically much broader than an exponential distribution, indicated by its larger CV (Eq. (71)); linking this with our condition on path length and time statistics $(\theta^{(cv)})^2 \ll \bar{\ell}$, we obtain a condition that shows how much coarse-graining is necessary to see significant memory effects in the statistics of path times (Eq. (72)).

Representing complex state spaces by simpler, coarse-grained representations has long been an implicit element of

stochastic models. In recent years, it has been explored in Markov models of molecular systems such as proteins.[1,2,16] Non-exponential effects may be important in these systems, especially if the coarse-grained networks are not very large. Indeed, non-exponential distributions of transition times have previously been found for both protein folding[17] and enzyme kinetics;[19] Reuveni et al.[20] showed that these memory effects could lead to qualitatively different properties of enzyme unbinding within the Michaelis-Menten framework. Our observations underscore the importance of going beyond characterizing such processes by single rates, which implicitly assumes an exponential distribution of times.

Besides waiting memory in the form of non-exponential time distributions, an additional form of memory induced by coarse-graining is in the jump process. For example, consider a triple-well potential coarse-grained into states A, B, and C. When the system crosses the barrier from A into B, it is much more likely to jump back into A rather than jump to C, since it begins much closer to A in the microscopic space. We can account for this in our framework by extending the state space to include not only the current state of the system (e.g., A, B, or C) but also the previous state; the jump process in this extended state space is once again Markovian (although the waiting time distributions remain non-exponential). We also note that this coarse-graining may require non-separable waiting time distributions $\psi(t|\sigma \to \sigma')$, since the distribution of times to return to A from B may be quite different from the distribution to reach C. Our framework can readily address this generalization (Appendix E). We look forward to studying the combined roles of jump and waiting memory in coarse-grained molecular models.

ACKNOWLEDGMENTS

APPENDIX A: ASYMPTOTIC FORM OF THE PATH LENGTH DISTRIBUTION

The path length distribution $\rho(\ell)$ is formally given by Eq. (12), which involves the sum of path probabilities $\mathcal{P}[\varphi]$ for all paths φ of length ℓ. More explicitly, we can write $\rho(\ell)$ using matrix elements of powers of the jump matrix $\mathbf{Q}$ and summing over all final states,

$$\rho(\ell) = \sum_{\sigma \in S_{\text{final}}} \langle \sigma | \mathbf{Q}^\ell | \pi_0 \rangle, \tag{A1}$$

where $|\pi_0\rangle = \sum_\sigma \pi_0(\sigma)|\sigma\rangle$ is the vector of initial state probabilities. We can decompose $\mathbf{Q}$ into its Jordan form

$$\mathbf{Q} = \mathbf{P}(\mathbf{D} + \mathbf{N})\mathbf{P}^{-1}, \tag{A2}$$

where $\mathbf{D}$ is a diagonal matrix with the eigenvalues of $\mathbf{Q}$, $\mathbf{N}$ is a nilpotent matrix, and $\mathbf{P}$ is an invertible matrix.[48] Powers of

$\mathbf{Q}$ are therefore

$$\mathbf{Q}^\ell = \mathbf{P}(\mathbf{D} + \mathbf{N})^\ell \mathbf{P}^{-1}$$
$$= \sum_{\ell'=0}^{\ell} \binom{\ell}{\ell'} \mathbf{P}\mathbf{D}^{\ell'}\mathbf{N}^{\ell-\ell'}\mathbf{P}^{-1}. \tag{A3}$$

If $\mathbf{Q}$ is exactly diagonalizable, then $\mathbf{N} = 0$, and so for large ℓ, the leading order term in Eq. (A1) is proportional to $q^\ell = e^{\ell \log q}$, where $q < 1$ is the largest eigenvalue of $\mathbf{Q}$. If $\mathbf{Q}$ is not diagonalizable, the leading term will still be proportional to $e^{\ell \log q}$, but may also include a polynomial factor in ℓ due to the binomial coefficient in Eq. (A3). However, the polynomial factor only contributes logarithmically to the exponent, i.e., $\ell^k e^{\ell \log q} = e^{\ell \log q + k \log \ell}$, and thus we can neglect it for large ℓ. Therefore, in general we have $\rho(\ell) \sim e^{\ell \log q}$ for large ℓ, and since this suggests that the mean path length must be $\bar{\ell} \sim -1/\log q$, we obtain Eq. (13).

APPENDIX B: EXACT RELATIONS BETWEEN PATH LENGTH AND TIME MOMENTS USING GENERATING FUNCTIONS

Here we derive exact relations between path length and time moments when all states have identical waiting time distributions: $\psi(t|\sigma) = \psi(t)$. We define the moment-generating function for the path length distribution

$$\tilde{\rho}(s) = \sum_{\ell=0}^{\infty} \rho(\ell) e^{s\ell}, \tag{B1}$$

so that the moments are

$$\langle \mathcal{L}^n \rangle = \tilde{\rho}^{(n)}(0), \tag{B2}$$

where the superscript denotes the derivative:

$$\tilde{\rho}^{(n)}(0) = \frac{d^n}{ds^n} \tilde{\rho}(s)\bigg|_{s=0}. \tag{B3}$$

The cumulant-generating function is therefore $\tilde{\rho}_c(s) = \log \tilde{\rho}(s)$ with

$$\langle \mathcal{L}^n \rangle_c = \tilde{\rho}_c^{(n)}(0). \tag{B4}$$

We similarly define the moment- and cumulant-generating functions for the waiting times:

$$\tilde{\psi}(s) = \int_0^\infty dt\, \psi(t) e^{st}, \qquad \theta^{(n)} = \tilde{\psi}^{(n)}(0),$$
$$\tilde{\psi}_c(s) = \log \tilde{\psi}(s), \qquad \theta_c^{(n)} = \tilde{\psi}_c^{(n)}(0). \tag{B5}$$

When the waiting time distributions are $\psi(t)$ for every state, the path time distribution is (Eqs. (7) and (14))

$$f(t) = \sum_{\ell=0}^{\infty} \rho(\ell) \int_0^\infty dt_0\, \psi(t_0) \int_0^\infty dt_1\, \psi(t_1) \cdots$$
$$\times \int_0^\infty dt_{\ell-1}\, \psi(t_{\ell-1})\, \delta\left(t - \sum_{i=0}^{\ell-1} t_i\right). \tag{B6}$$

Therefore, the moment-generating function for path time is

$$
\begin{aligned}
\tilde{f}(s) &= \int_0^\infty dt \, f(t) \, e^{st} \\
&= \sum_{\ell=0}^\infty \rho(\ell)(\tilde{\psi}(s))^\ell \\
&= \sum_{\ell=0}^\infty \rho(\ell)e^{\ell \log \tilde{\psi}(s)} \\
&= \tilde{\rho}\left(\tilde{\psi}_{\mathrm{c}}(s)\right),
\end{aligned}
\tag{B7}
$$

while the cumulant-generating function for path time is

$$
\begin{aligned}
\tilde{f}_{\mathrm{c}}(s) &= \log \tilde{f}(s) \\
&= \log \tilde{\rho}\left(\tilde{\psi}_{\mathrm{c}}(s)\right) \\
&= \tilde{\rho}_{\mathrm{c}}\left(\tilde{\psi}_{\mathrm{c}}(s)\right).
\end{aligned}
\tag{B8}
$$

We can obtain moments and cumulants of path time by taking derivatives of its generating functions:

$$
\left\langle \mathcal{T}^{(n)} \right\rangle = \tilde{f}^{(n)}(0), \quad \left\langle \mathcal{T}^{(n)} \right\rangle_{\mathrm{c}} = \tilde{f}_{\mathrm{c}}^{(n)}(0).
\tag{B9}
$$

To express these in terms of the length and waiting time moments, we use Faà di Bruno's formula for derivatives of composite functions:[37]

$$
\begin{aligned}
\frac{d^n}{ds^n} g(h(s)) &= \sum_{k=1}^n g^{(k)}(h(s)) B_{n,k}(h^{(1)}(s), h^{(2)}(s), \dots, \\
& \qquad h^{(n-k+1)}(s)),
\end{aligned}
\tag{B10}
$$

where $B_{n,k}$ are the partial Bell polynomials and superscripts again denote derivatives. Thus, the path time moments are

$$
\begin{aligned}
\left\langle \mathcal{T}^{(n)} \right\rangle &= \tilde{f}^{(n)}(0) \\
&= \frac{d^n}{ds^n} \tilde{\rho}\left(\tilde{\psi}_{\mathrm{c}}(s)\right)\bigg|_{s=0} \\
&= \sum_{k=1}^n \tilde{\rho}^{(k)}(\tilde{\psi}_{\mathrm{c}}(0)) \\
& \quad \times B_{n,k}\left(\tilde{\psi}_{\mathrm{c}}^{(1)}(0), \tilde{\psi}_{\mathrm{c}}^{(2)}(0), \dots, \tilde{\psi}_{\mathrm{c}}^{(n-k+1)}(0)\right) \\
&= \sum_{k=1}^n \left\langle \mathcal{L}^n \right\rangle B_{n,k}\left(\theta_{\mathrm{c}}^{(1)}, \theta_{\mathrm{c}}^{(2)}, \dots, \theta_{\mathrm{c}}^{(n-k+1)}\right).
\end{aligned}
\tag{B11}
$$

This proves Eq. (24). We can similarly obtain the path time cumulants:

$$
\begin{aligned}
\left\langle \mathcal{T}^{(n)} \right\rangle_{\mathrm{c}} &= \tilde{f}_{\mathrm{c}}^{(n)}(0) \\
&= \frac{d^n}{ds^n} \tilde{\rho}_{\mathrm{c}}\left(\tilde{\psi}_{\mathrm{c}}(s)\right)\bigg|_{s=0} \\
&= \sum_{k=1}^n \tilde{\rho}_{\mathrm{c}}^{(k)}(\tilde{\psi}_{\mathrm{c}}(0)) \\
& \quad \times B_{n,k}\left(\tilde{\psi}_{\mathrm{c}}^{(1)}(0), \tilde{\psi}_{\mathrm{c}}^{(2)}(0), \dots, \tilde{\psi}_{\mathrm{c}}^{(n-k+1)}(0)\right) \\
&= \sum_{k=1}^n \left\langle \mathcal{L}^n \right\rangle_{\mathrm{c}} B_{n,k}\left(\theta_{\mathrm{c}}^{(1)}, \theta_{\mathrm{c}}^{(2)}, \dots, \theta_{\mathrm{c}}^{(n-k+1)}\right).
\end{aligned}
\tag{B12}
$$

APPENDIX C: APPROXIMATE RELATIONS BETWEEN PATH LENGTH AND TIME MOMENTS USING THE CENTRAL LIMIT THEOREM

Here we use the central limit theorem to obtain approximate relations between path length and time moments when all states have identical waiting time distributions: $\psi(t|\sigma) = \psi(t)$. Equation (B6) gives the general relation between the length and time distributions in this case, where the nested integrals represent the probability distribution of the sum of the waiting times. For sufficiently long paths, this distribution will be approximately Gaussian

$$
\frac{1}{\sqrt{2\pi \ell \theta_{\mathrm{c}}^{(2)}}} \exp\left(-\frac{\left(t - \ell \theta_{\mathrm{c}}^{(1)}\right)^2}{2\ell \theta_{\mathrm{c}}^{(2)}}\right),
\tag{C1}
$$

and hence,

$$
f(t) \approx \sum_{\ell=0}^\infty \rho(\ell) \frac{1}{\sqrt{2\pi \ell \theta_{\mathrm{c}}^{(2)}}} \exp\left(-\frac{\left(t - \ell \theta_{\mathrm{c}}^{(1)}\right)^2}{2\ell \theta_{\mathrm{c}}^{(2)}}\right).
\tag{C2}
$$

From this, we can obtain approximate relations between the moments. For example, the first two path time moments are

$$
\begin{aligned}
\left\langle \mathcal{T}^{(1)} \right\rangle &\approx \theta_{\mathrm{c}}^{(1)} \left\langle \mathcal{L} \right\rangle, \\
\left\langle \mathcal{T}^{(2)} \right\rangle &\approx (\theta_{\mathrm{c}}^{(1)})^2 \left\langle \mathcal{L}^2 \right\rangle + \theta_{\mathrm{c}}^{(2)} \left\langle \mathcal{L} \right\rangle,
\end{aligned}
\tag{C3}
$$

which are in fact identical to the exact result (Eqs. (24) and (25)) as expected.

APPENDIX D: PATH LENGTHS AND TIMES WITH HOMOGENEOUS EXPONENTIAL WAITING TIME DISTRIBUTIONS

When every state has the same exponential waiting time distribution $\psi(t) = \theta^{-1} e^{-t/\theta}$, the sum of the ℓ waiting times has an Erlang distribution,

$$
\begin{aligned}
& \int_0^\infty dt_0 \, \frac{1}{\theta} e^{-t_0/\theta} \int_0^\infty dt_1 \, \frac{1}{\theta} e^{-t_1/\theta} \cdots \\
& \quad \times \int_0^\infty dt_{\ell-1} \, \frac{1}{\theta} e^{-t_{\ell-1}/\theta} \, \delta\left(t - \sum_{i=0}^{\ell-1} t_i\right) \\
& = \frac{(t/\theta)^{\ell-1}}{(\ell-1)!\theta} e^{-t/\theta}.
\end{aligned}
\tag{D1}
$$

Thus, the total path time distribution is (using Eq. (B6))

$$
f(t) = \frac{1}{\theta} e^{-t/\theta} \sum_{\ell=0}^\infty \rho(\ell) \frac{(t/\theta)^{\ell-1}}{(\ell-1)!}.
\tag{D2}
$$

In this case, we can determine the complete distribution of times $f(t)$ given the complete distribution of lengths $\rho(\ell)$. We

can directly calculate the path time moments to be

$$
\begin{aligned}
\left\langle \mathcal{T}^{(n)} \right\rangle &= \int_0^\infty dt\, f(t)\, t^n \\
&= \sum_{\ell=0}^\infty \rho(\ell) \int_0^\infty dt\, \frac{(t/\theta)^{\ell-1}}{\theta(\ell-1)!} e^{-t/\theta} t^n \\
&= \sum_{\ell=0}^\infty \rho(\ell) \theta^n \frac{(n+\ell-1)!}{(\ell-1)!} \\
&= \theta^n \left\langle \mathcal{L}(\mathcal{L}+1)\cdots(\mathcal{L}+n-1) \right\rangle \\
&= \theta^n \sum_{k=1}^n \left\langle \mathcal{L}^k \right\rangle |s_{n,k}|,
\end{aligned}
\tag{D3}
$$

where $|s_{n,k}|$ are the unsigned Stirling numbers of the first kind.[37] The first few moments are

$$
\begin{aligned}
\left\langle \mathcal{T}^{(1)} \right\rangle &= \theta \left\langle \mathcal{L} \right\rangle, \\
\left\langle \mathcal{T}^{(2)} \right\rangle &= \theta^2 \left(\left\langle \mathcal{L}^2 \right\rangle + \left\langle \mathcal{L} \right\rangle \right), \\
\left\langle \mathcal{T}^{(3)} \right\rangle &= \theta^3 \left(\left\langle \mathcal{L}^3 \right\rangle + 3\left\langle \mathcal{L}^2 \right\rangle + 2\left\langle \mathcal{L} \right\rangle \right).
\end{aligned}
\tag{D4}
$$

This is consistent with the general result in Eq. (24) since $\theta_c^{(j)} = (j-1)!\,\theta^j$ for an exponential distribution and $B_{n,k}(0!,1!,\ldots,(n-k)!) = |s_{n,k}|$.[37]

APPENDIX E: PROOF OF RECURSION RELATIONS FOR MOMENT MATRICES

We now show that the $\mathbf{T}_\ell^{(n)}$ matrices generated by the recursion relation of Eq. (34) indeed calculate the path time moments according to Eq. (33). We first successively apply the recursion relation to expand the ℓth-order matrix in terms of lower-order matrices:

$$
\begin{aligned}
\mathbf{T}_\ell^{(n)} &= \mathbf{Q} \sum_{j_{\ell-1}=0}^n \binom{n}{j_{\ell-1}} \boldsymbol{\Theta}^{(j_{\ell-1})} \mathbf{T}_{\ell-1}^{(n-j_{\ell-1})} \\
&= \mathbf{Q} \sum_{j_{\ell-1}=0}^n \binom{n}{j_{\ell-1}} \boldsymbol{\Theta}^{(j_{\ell-1})} \mathbf{Q} \sum_{j_{\ell-2}=0}^{n-j_{\ell-1}} \binom{n-j_{\ell-1}}{j_{\ell-2}} \boldsymbol{\Theta}^{(j_{\ell-2})} \mathbf{T}_{\ell-2}^{(n-j_{\ell-1}-j_{\ell-2})} \\
&\;\;\vdots \\
&= \sum_{j_{\ell-1}=0}^n \sum_{j_{\ell-2}=0}^{n-j_{\ell-1}} \cdots \sum_{j_0=0}^{n-j_{\ell-1}-j_{\ell-2}-\cdots-j_1} \binom{n}{j_{\ell-1}} \binom{n-j_{\ell-1}}{j_{\ell-2}} \cdots \binom{n-j_{\ell-1}-j_{\ell-2}\cdots-j_1}{j_0} \\
&\quad \times \mathbf{Q}\boldsymbol{\Theta}^{(j_{\ell-1})} \mathbf{Q}\boldsymbol{\Theta}^{(j_{\ell-2})} \cdots \mathbf{Q}\boldsymbol{\Theta}^{(j_0)} \mathbf{T}_0^{(n-j_{\ell-1}-j_{\ell-2}-\cdots-j_0)} \\
&= \sum_{j_0,j_1,\ldots,j_{\ell-1}} \binom{n}{j_0,j_1,\ldots,j_{\ell-1}} \mathbf{Q}\boldsymbol{\Theta}^{(j_{\ell-1})} \mathbf{Q}\boldsymbol{\Theta}^{(j_{\ell-2})} \cdots \mathbf{Q}\boldsymbol{\Theta}^{(j_0)},
\end{aligned}
\tag{E1}
$$

where we have invoked the initial condition $\mathbf{T}_0^{(n-j_{\ell-1}-j_{\ell-2}-\cdots-j_0)} = \delta_{0,n-j_{\ell-1}-j_{\ell-2}-\cdots-j_0}\mathbf{1}$ from Eq. (31) to obtain the multinomial sum (recall that each summation in the multinomial sum is from 0 to n subject to the constraint $j_0 + j_1 + \cdots + j_{\ell-1} = n$). Now we take the matrix element of $\mathbf{T}_\ell^{(n)}$ for the initial distribution $|\pi_0\rangle = \sum_\sigma \pi_0(\sigma)|\sigma\rangle$ and $\sigma_\ell \in S_{\text{final}}$ and insert identities of the form $\sum_\sigma |\sigma\rangle\langle\sigma|$ to obtain

$$
\begin{aligned}
\langle \sigma_\ell | \mathbf{T}_n^{(\ell)} | \pi_0 \rangle &= \sum_{j_0,j_1,\ldots,j_{\ell-1}} \binom{n}{j_0,j_1,\ldots,j_{\ell-1}} \\
&\quad \times \sum_{\sigma_0,\sigma_1,\ldots,\sigma_{\ell-1}} \langle\sigma_\ell|\mathbf{Q}\boldsymbol{\Theta}^{(j_{\ell-1})}|\sigma_{\ell-1}\rangle\langle\sigma_{\ell-1}|\mathbf{Q}\boldsymbol{\Theta}^{(j_{\ell-2})}|\sigma_{\ell-2}\rangle \cdots \langle\sigma_1|\mathbf{Q}\boldsymbol{\Theta}^{(j_0)}|\sigma_0\rangle \pi_0(\sigma_0) \\
&= \sum_{\sigma_0,\sigma_1,\ldots,\sigma_{\ell-1}} \langle\sigma_\ell|\mathbf{Q}|\sigma_{\ell-1}\rangle\langle\sigma_{\ell-1}|\mathbf{Q}|\sigma_{\ell-2}\rangle \cdots \langle\sigma_1|\mathbf{Q}|\sigma_0\rangle \pi_0(\sigma_0) \\
&\quad \times \sum_{j_0,j_1,\ldots,j_{\ell-1}} \binom{n}{j_0,j_1,\ldots,j_{\ell-1}} \theta^{(j_0)}(\sigma_0)\theta^{(j_1)}(\sigma_1)\cdots\theta^{(j_{\ell-1})}(\sigma_{\ell-1}).
\end{aligned}
\tag{E2}
$$

Next, we sum over final states σ_ℓ and path lengths ℓ to obtain

$$
\begin{aligned}
\sum_{\ell=0}^\infty \sum_{\sigma_\ell \in S_{\text{final}}} \langle\sigma|\mathbf{T}_\ell^{(n)}|\pi_0\rangle &= \sum_{\ell=0}^\infty \sum_{\sigma_\ell \in S_{\text{final}}} \sum_{\sigma_0,\sigma_1,\ldots,\sigma_{\ell-1}} \langle\sigma_\ell|\mathbf{Q}|\sigma_{\ell-1}\rangle\langle\sigma_{\ell-1}|\mathbf{Q}|\sigma_{\ell-2}\rangle \cdots \langle\sigma_1|\mathbf{Q}|\sigma_0\rangle \pi_0(\sigma_0) \\
&\quad \times \sum_{j_0,j_1,\ldots,j_{\ell-1}} \binom{n}{j_0,j_1,\ldots,j_{\ell-1}} \theta^{(j_0)}(\sigma_0)\theta^{(j_1)}(\sigma_1)\cdots\theta^{(j_{\ell-1})}(\sigma_{\ell-1}).
\end{aligned}
\tag{E3}
$$

Substituting $\mathcal{P}[\varphi]$ (Eq. (8)), the time moment functional $\mathcal{T}^{(n)}[\varphi]$ (Eq. (16)), and the sum over paths

$$\sum_{\varphi} = \sum_{\ell=0}^{\infty} \sum_{\sigma_\ell \in S_{\text{final}}} \sum_{\sigma_0, \sigma_1, \ldots, \sigma_{\ell-1}} \tag{E4}$$

into Eq. (E3), we finally obtain Eq. (33):

$$\sum_{\ell=0}^{\infty} \sum_{\sigma_\ell \in S_{\text{final}}} \langle \sigma | \mathbf{T}_\ell^{(n)} | \pi_0 \rangle = \sum_{\varphi} \mathcal{P}[\varphi] \mathcal{T}^{(n)}[\varphi]$$

$$= \left\langle \mathcal{T}^{(n)} \right\rangle. \tag{E5}$$

We now show that the recursion relation of Eq. (40) is correct for any functional of the form in Eq. (22); this will prove the action recursion relation (Eq. (37)) as a special case. As in Eq. (E1), successive applications of the $\mathbf{U}_\ell^{(n)}$ recursion relation yield

$$\mathbf{U}_\ell^{(n)} = \sum_{j_0, j_1, \ldots, j_{\ell-1}} \binom{n}{j_0, j_1, \ldots, j_{\ell-1}} \boldsymbol{\Omega}^{(j_{\ell-1})} \boldsymbol{\Omega}^{(j_{\ell-2})} \cdots \boldsymbol{\Omega}^{(j_0)}. \tag{E6}$$

After inserting identities and using the definition $\langle \sigma' | \boldsymbol{\Omega}^{(j)} | \sigma \rangle = \langle \sigma' | \mathbf{Q} | \sigma \rangle (U(\sigma', \sigma))^j$, we obtain

$$\langle \sigma_\ell | \mathbf{U}_n^{(\ell)} | \pi_0 \rangle = \sum_{\sigma_0, \sigma_1, \ldots, \sigma_{\ell-1}} \langle \sigma_\ell | \mathbf{Q} | \sigma_{\ell-1} \rangle \langle \sigma_{\ell-1} | \mathbf{Q} | \sigma_{\ell-2} \rangle \cdots \langle \sigma_1 | \mathbf{Q} | \sigma_0 \rangle \pi_0(\sigma_0)$$

$$\times \sum_{j_0, j_1, \ldots, j_{\ell-1}} \binom{n}{j_0, j_1, \ldots, j_{\ell-1}} (U(\sigma_\ell, \sigma_{\ell-1}))^{j_{\ell-1}} (U(\sigma_{\ell-1}, \sigma_{\ell-2}))^{j_{\ell-2}} \cdots (U(\sigma_1, \sigma_0))^{j_0}. \tag{E7}$$

Just as in Eqs. (E3) and (E5) for time moments, we can then sum over final states and path lengths to show that the $\mathbf{U}_\ell^{(n)}$ matrices are related to the path ensemble averages via Eq. (39). Finally, we explain how to use the generalized recursion relation in Eq. (40) to calculate path time moments when the waiting time distributions are non-separable. In this case, we define $\langle \sigma' | \boldsymbol{\Omega}^{(j)} | \sigma \rangle = \langle \sigma' | \mathbf{Q} | \sigma \rangle \theta^{(j)}(\sigma \to \sigma')$, where $\theta^{(j)}(\sigma \to \sigma')$ is the jth moment of the non-separable waiting time distribution $\psi(t | \sigma \to \sigma')$. We can then use the recursion relation of Eq. (40) to calculate the time moments $\mathbf{U}_\ell^{(n)} = \mathbf{T}_\ell^{(n)}$.

[1]F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, "Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations," Proc. Natl. Acad. Sci. U. S. A. **106**, 19011–19016 (2009).

[2]T. J. Lane, G. R. Bowman, K. Beauchamp, V. A. Voelz, and V. S. Pande, "Markov state model reveals folding and functional dynamics in ultra-long MD trajectories," J. Am. Chem. Soc. **133**, 18413–18419 (2011).

[3]D. ben-Avraham and S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge University Press, Cambridge, 2000).

[4]S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, 2001).

[5]D. M. Weinreich, N. F. Delaney, M. A. DePristo, and D. L. Hartl, "Darwinian evolution can follow only very few mutational paths to fitter proteins," Science **312**, 111–114 (2006).

[6]M. Manhart and A. V. Morozov, "Protein folding and binding can emerge as evolutionary spandrels through structural coupling," Proc. Natl. Acad. Sci. U. S. A. **112**, 1797–1802 (2015).

[7]T. Enver, M. Pera, C. Peterson, and P. W. Andrews, "Stem cell states, fates, and the rules of attraction," Cell Stem Cell **4**, 387 (2009).

[8]A. H. Lang, H. Li, J. J. Collins, and P. Mehta, "Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes," PLoS Comput. Biol. **10**, e1003734 (2014).

[9]R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," Rev. Mod. Phys. **74**, 47–97 (2002).

[10]L. K. Gallos, C. Song, S. Havlin, and H. A. Makse, "Scaling theory of transport in complex biological networks," Proc. Natl. Acad. Sci. U. S. A. **104**, 7746–7751 (2007).

[11]S. Condamin, O. Bénichou, V. Tejedor, R. Voituriez, and J. Klafter, "First-passage times in complex scale-invariant media," Nature **450**, 77–80 (2007).

[12]E. W. Montroll and G. H. Weiss, "Random walks on lattices. II," J. Math. Phys. **6**, 167–181 (1965).

[13]G. H. Weiss, *Aspects and Applications of the Random Walk* (North Holland, Amsterdam, 1994).

[14]E. M. Bollt and D. ben-Avraham, "What is special about diffusion on scale-free nets?," New J. Phys. **7**, 26–47 (2005).

[15]P. Metzner, C. Schütte, and E. Vanden-Eijnden, "Transition path theory for Markov jump processes," Multiscale Model. Simul. **7**, 1192–1219 (2009).

[16]J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, "Markov models of molecular kinetics: Generation and validation," J. Chem. Phys. **134**, 174105 (2011).

[17]J. Sabelko, J. Ervin, and M. Gruebele, "Observation of strange kinetics in protein folding," Proc. Natl. Acad. Sci. U. S. A. **96**, 6031–6036 (1999).

[18]H. Yang, G. Luo, P. Karnchanaphanurach, T.-M. Louie, I. Rech, S. Cova, L. Xun, and X. S. Xie, "Protein conformational dynamics probed by single-molecule electron transfer," Science **302**, 262–266 (2003).

[19]O. Flomenbom, K. Velonia, D. Loos, S. Masuo, M. Cotlet, Y. Engelborghs, J. Hofkens, A. E. Rowan, R. J. M. Nolte, M. V. der Auweraer, F. C. de Schryver, and J. Klafter, "Stretched exponential decay and correlations in the catalytic activity of fluctuating single lipase molecules," Proc. Natl. Acad. Sci. U. S. A. **102**, 2368–2372 (2005).

[20]S. Reuveni, M. Urbakh, and J. Klafter, "Role of substrate unbinding in Michaelis–Menten enzymatic reactions," Proc. Natl. Acad. Sci. U. S. A. **111**, 4391–4396 (2014).

[21]I. A. Campbell, J.-M. Flesselles, R. Jullien, and R. Botet, "Nonexponential relaxation in spin glasses and glassy systems," Phys. Rev. B **37**, 3825–3828 (1988).

[22]L. Angelani, G. Parisi, G. Ruocco, and G. Viliani, "Connected network of minima as a model glass: Long time dynamics," Phys. Rev. Lett. **81**, 4648–4651 (1998).

[23]J. J. Hunter, "On the moments of Markov renewal processes," Adv. Appl. Probab. **1**, 188–210 (1969).

[24]D. D. Yao, "First-passage-time moments of Markov processes," J. Appl. Probab. **22**, 939–945 (1985).

[25]P. G. Harrison and W. J. Knottenbelt, "Passage time distributions in large Markov chains," SIGMETRICS Perform. Eval. Rev. **30**, 77–85 (2002).

[26]O. Flomenbom and J. Klafter, "Closed-form solutions for continuous time random walks on finite chains," Phys. Rev. Lett. **95**, 098105 (2005).

[27]O. Flomenbom and R. J. Silbey, "Properties of the generalized master equation: Green's functions and probability density functions in the path representation," J. Chem. Phys. **127**, 034103 (2007).

[28]O. Flomenbom and R. J. Silbey, "Path-probability density functions for semi-Markovian random walks," Phys. Rev. E **76**, 041101 (2007).

[29]B. Harland and S. X. Sun, "Path ensembles and path sampling in nonequilibrium stochastic systems," J. Chem. Phys. **127**, 104103 (2007).

[30]J. W. Haus and K. W. Kehr, "Diffusion in regular and disordered lattices," Phys. Rep. **150**, 263–406 (1987).

[31]S. X. Sun, "Path summation formulation of the master equation," Phys. Rev. Lett. **96**, 210602 (2006).

[32]M. Manhart and A. V. Morozov, "Path-based approach to random walks on networks characterizes how proteins evolve new functions," Phys. Rev. Lett. **111**, 088102 (2013).

[33]M. Manhart and A. V. Morozov, "Statistical physics of evolutionary trajectories on fitness landscapes," in *First-Passage Phenomena and their Applications*, edited by R. Metzler, G. Oshanin, and S. Redner (World Scientific, Singapore, 2014).

[34]A. A. Filyukov and V. Y. Karpov, "Method of the most probable path of evolution in the theory of stationary irreversible processes," J. Eng. Phys. **13**, 416–419 (1967).

[35]V. Mustonen and M. Lässig, "Fitness flux and ubiquity of adaptive evolution," Proc. Natl. Acad. Sci. U. S. A. **107**, 4248–4253 (2010).

[36]J. Wang, L. Xu, and E. Wang, "Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations," Proc. Natl. Acad. Sci. U. S. A. **105**, 12271–12276 (2008).

[37]L. Comtet, *Advanced Combinatorics: The Art of Finite and Infinite Expressions* (D. Reidel, Dordrecht, 1974).

[38]J. M. Yeomans, *Statistical Mechanics of Phase Transitions* (Oxford University Press, Oxford, 1992).

[39]I. Majid, D. ben Avraham, S. Havlin, and H. E. Stanley, "Exact-enumeration approach to random walks on percolation clusters in two dimensions," Phys. Rev. B **30**, 1626–1628 (1984).

[40]E. Jones, T. Oliphant, P. Peterson *et al.*, SciPy: Open source scientific tools for Python, 2001, online accessed May 15, 2015.

[41]N. S. Goel and N. Richter-Dyn, *Stochastic Models in Biology* (Academic Press, New York, 1974).

[42]W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992).

[43]A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics: Volume 1*, 6th ed. (Wiley, New York, 1994).

[44]G. Bel, B. Munsky, and I. Nemenman, "The simplicity of completion time distributions for common complex biochemical processes," Phys. Biol. **7**, 016003 (2010).

[45]G. H. Weiss and S. Havlin, "Use of comb-like models to mimic anomalous diffusion on fractal structures," Philos. Mag. B **56**, 941–947 (1987).

[46]A. Iomin, "Toy model of fractional transport of cancer cells due to self-entrapping," Phys. Rev. E **73**, 061918 (2006).

[47]P. Argyrakis, A. Milchev, V. Pereyra, and K. W. Kehr, "Dependence of the diffusion coefficient on the energy distribution of random barriers," Phys. Rev. E **52**, 3623–3631 (1995).

[48]C. W. Curtis, *Linear Algebra: An Introductory Approach* (Springer, New York, 1984).